THE PEDAGOGICAL SEMINARY AND

JOURNAL OF GENETIC PSYCHOLOGY

Child Behavior, Animal Behavior, and Comparative Psychology

EDITED BY
CARL MURCHISON

John E. Anderson
University of Minnesota

Charlotte Bühler
Universität, Wien

William H. Burnham
Clark University

Cyril Burt
University of London

Leonard Carmichael
University of Rochester

Ed. Claparède
Université de Genève

Edmund S. Conklin
Indiana University

Sante De Sanctis
R. Università di Roma

Arnold Gesell
Yale University

William Healy
Judge Baker Foundation, Boston

Leta S. Hollingworth
Teachers College, Columbia University

Buford Johnson
The Johns Hopkins University

Harold E. Jones
University of California

Truman L. Kelley
Harvard University

Yoshihide Kubo
Hiroshima Normal College

K. S. Lashley
Harvard University

A. R. Luria
Medico-biological Institute, Moscow

Toshio Nogami
Kyoto Imperial University

Henri Piéron
Université de Paris

Calvin P. Stone
Stanford University

Lewis M. Terman
Stanford University

Godfrey Thomson
University of Edinburgh

E. L. Thorndike
Teachers College, Columbia University

C. J. Warden
Columbia University

John B. Watson
New York City

Helen Thompson Woolley
Teachers College, Columbia University

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold under the restrictions of a bilateral contract, and may not be resold for a period of five years from date of publication.

VOLUME 52

1938

Copyright, 1938, by The Journal Press
Published quarterly by The Journal Press
Provincetown, Massachusetts, U. S. A.

AUTHOR INDEX

Ball, Josephine 439
Ball, Rachel Stutsman 79
Banker, Mary H 347
Beckman, R O. 203
Benton, Arthur L 395
Bornemeier, Russell W. 165
Carmichael, Leonard 159
Clarke, Frances M. 443
Drew, G. C 257
Durea, Mervin A. 269
Farnsworth, Paul R 3
Fialkin, H N 203
Forbes, Marie L. H 247, 147
Ghiselli, E E 151
Goodenough, Florence L 251, 333
Halstead, Ward 227
Hertz, Marguerite R. 15
Honzik, Marjorie Pyles 285
Hunt, Edna B 143
Hunt, William A. 443
Kantor, Robert 211
Kohn, Harold A 433
Kounin, J S 375
Krechevsky, I 241
Lichtenstein, Arthur 407
McGehee, William 361
Morgan, John J. B 347
Nelson, Virginia Lafayette 303, 327
Richards, T. W. 303, 327
Roberts, Katherine Elliott 79
Snygg, Donald 401
Wagner, Isabel F 65, 233, 235
Walton, William E. 165
Warden, Carl J 211
Wentworth, Kenneth L. 429
Wheeler, Lester R. 183
Winslow, Charles Nelson 211, 425
Yacorzynski, George 227

TABLE OF CONTENTS

A genetic study of the Bernreuter personality inventory . . . 3
By Paul R. Farnsworth

Scoring the Rorschach ink-blot test 15
By Marguerite R. Hertz

The body jerk of the neonate 65
By Isabelle F. Wagner

A study of personality in young children by means of a series of rating scales 79
By Katherine Elliott Roberts and Rachel Stutsman Ball

The relationship between the superior colliculus and the striate area in brightness discrimination 151
By E. E. Ghiselli

Learning which modifies an animal's subsequent capacity for learning 159
By Leonard Carmichael

Further evidence of color discrimination in rodents . . 165
By William E. Walton and Russell W. Bornemeier

A study of the relative difficulty of a primary reading vocabulary 183
By Lester R. Wheeler

The influence of month of birth on the intelligence test scores of adults 203
By H. N. Fialkin and R. O. Beckman

An investigation of conditioning in cats to multiple stimuli . . 211
By Charles Nelson Winslow, Robert Kantor, and Carl J. Warden

SHORT ARTICLES AND NOTES 227

A jumping method for establishing differential responses in pigeons (Ward Halstead and George Yacorzynski) . . 227

A note on the hiccough of the neonate (Isabelle F. Wagner) . 233

The sleeping posture of the neonate (Isabelle F. Wagner) . 235

A note on the perception of linear gestalten in the rat (I. Krechevsky) 241

APPARATUS 247

The block printing stick as a device for registering choice (Marie L. H. Forbes) . . . 247

BOOKS 251

Lois Barclay Murphy, *Social behavior and child personality* (Florence L. Goodenough) 251

The function of punishment in learning 257
By G. C. Drew

Personality characteristics of juvenile offenders in relation to degree of delinquency 269
By Mervin A. Durea

The constancy of mental test performance during the preschool period 285
By Marjorie Pyles Honzik

Studies in mental development: I. Performance on Gesell items at six months and its predictive value for performance on mental tests at two and three years 303
By Virginia Lafayette Nelson and T. W. Richards

Studies in mental development: II. Analysis of abilities tested at the age of six months by the Gesell schedule 327
By T. W. Richards and Virginia Lafayette Nelson

The use of pronouns by young children: A note on the development of self-awareness 333
By Florence L. Goodenough

The relation of mental stamina to parental protection 347
By John J. B. Morgan and Mary H. Banker

The free word association of elementary school children: II. Verbal responses 361
By William McGehee

Laterality in monkeys 375
By J. S. Kounin

Performance of school children on the revised Stanford-Binet and the Kent E-G-Y test 393
By Arthur L. Benton

The relation between the intelligence of mothers and of their children living in foster homes 401
By Donald Snygg

An investigation of reading retardation 407
By Arthur Lichtenstein

SHORT ARTICLES AND NOTES 425

- Observations of dominance-subordination in cats (Charles Nelson Winslow) 425
- The effect of early reaches on handedness in the rat. A preliminary study (Kenneth L. Wentworth) 429
- Achievement and intelligence examinations correlated with each other and with teacher's rankings (Harold A. Kohn) . . 433
- A case of apparent imitation in a monkey (Josephine Ball) . . 439
- The startle pattern in infants in response to non-auditory stimuli (William A. Hunt, Frances M. Clarke, and Edna B. Hunt) . 443

APPARATUS 447

- A combined block printing cell and tape holder (Marie L. H. Forbes) 447

BOOKS 449

THESE JOURNALS WILL BUY BOOK REVIEWS . . 449

BOOKS RECENTLY RECEIVED . . . 453

$7.00 per volume
Single numbers $4.00

QUARTERLY
Two volumes per year

March, 1938
Volume 52, First Half

Founded by G. Stanley Hall in 1891

THE PEDAGOGICAL SEMINARY AND JOURNAL OF GENETIC PSYCHOLOGY

Child Behavior, Animal Behavior, and Comparative Psychology

EDITED BY
CARL MURCHISON

John E. Anderson
University of Minnesota
Charlotte Bühler
Universität, Wien
William H. Burnham
Clark University
Cyril Burt
University of London
Leonard Carmichael
University of Rochester
Ed. Claparède
Université de Genève
Edmund S. Conklin
Indiana University
Sante De Sanctis
R. Università di Roma
Arnold Gesell
Yale University
William Healy
Judge Baker Foundation,
Boston
Leta S. Hollingworth
Teachers College,
Columbia University
Buford Johnson
The Johns Hopkins University
Harold E. Jones
University of California
Truman L. Kelley
Harvard University

Yoshihide Kubo
Hiroshima Normal College
K. S. Lashley
Harvard University
A. R. Luria
Medico-Biological Institute,
Moscow
Toshio Nogami
Kyoto Imperial University
Henri Piéron
Université de Paris
William Stern
Hamburgische Universität
Calvin P. Stone
Stanford University
Lewis M. Terman
Stanford University
Godfrey Thomson
University of Edinburgh
E. L. Thorndike
Teachers College,
Columbia University
C. J. Warden
Columbia University
John B. Watson
New York City
Helen Thompson Woolley
Teachers College,
Columbia University

Copyright, 1938, by The Journal Press
Provincetown, Massachusetts

Entered as second-class matter August 3, 1897, at the post-office at
Worcester, Mass., under Act of March 3, 1879
Reentered as second-class matter May 11, 1927, at the post-office at
Provincetown, Mass., under the Act of March 3, 1879

The Journal of Psychology

EDITED BY

Carl Murchison

The *Journal of Psychology* was founded at the request of many psychologists throughout the world for the purpose of establishing a medium in which publication would be immediate, and by which other publication experiments may be carried out. Beginning with January, 1938, the following methods and forms will be in force:

(1) The size of each volume will be 600-650 pages, and the subscription price will be $7 per volume, two volumes per year.

(2) The format of each volume will be such that each subscriber may cut apart the assembled volume into its component articles.

(3) Except through the fault of the author, the lag in publication should not be more than three or four weeks.

(4) The author will receive 200 reprints of his paper, and the subscribers will receive the accumulated publications bound into half volume units. The second half of each volume will contain the title page and table of contents for that volume. Each half volume unit will be mailed to the subscriber at its completion.

(5) Each author will be charged the actual manufacturing costs of any zinc or half-tone cuts.

(6) The publication charges will be as follows:

4 pages (in signature)	$14
[illegible]	18
[illegible]	24
[illegible]	32
[illegible]	40
[illegible]	48
[illegible]	56
[illegible]	64
[illegible]	72
[illegible]	90

These charges are tentative and will be adjusted as justified by the experience.

(7) The accumulated units will be mailed to subscribers quarterly (January, April, July, and October). The units may vary in size, to accommodate publication needs.

Beginning with January, 1938, this journal will contain twelve hundred pages (two volumes) annually. Per annum $14.00; per volume $7.00; per half volume $3.50; single copies [illegible] per number, plus transportation.

THE JOURNAL PRESS
2 Commercial Street
Provincetown, Massachusetts
U. S. A.

$7.00 per volume
Single numbers $4.00

QUARTERLY
Two volumes per year

March, 1938
Volume 52, First Half

Founded by G. Stanley Hall in 1891

THE PEDAGOGICAL SEMINARY AND
JOURNAL OF GENETIC PSYCHOLOGY

Child Behavior, Animal Behavior, and Comparative Psychology

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold under the restrictions of a bilateral contract, and may not be resold for a period of five years from date of publication

MARCH, 1938

(Manuscripts are printed in the order of final acceptance)

A genetic study of the Bernreuter personality inventory . . . 3
By Paul R Farnsworth

Scoring the Rorschach ink-blot test . . . 15
By Marguerite R Hertz

The body jerk of the neonate . . . 65
By Isabelle F Wagner

A study of personality in young children by means of a series of rating scales . . . 79
By Katherine Elliott Roberts and Rachel Stutsman Ball

(OVER)

Copyright, 1938, by The Journal Press

Provincetown, Massachusetts
Entered as second-class matter August 3, 1897, at the post-office at Worcester, Mass, under the act of March 3, 1879
Reentered as second-class matter May 11, 1937, at the post-office at Provincetown, Mass., under the Act of March 3, 1879

The relationship between the superior colliculus and the striate area in brightness discrimination 151
By E. E. Ghiselli

Learning which modifies an animal's subsequent capacity for learning 159
By Leonard Carmichael

Further evidence of color discrimination in rodents 165
By William E. Walton and Russell W. Bornemeier

A study of the relative difficulty of a primary reading vocabulary 183
By Lester R. Wheeler

The influence of month of birth on the intelligence test scores of adults 203
By H. N. Fialkin and R. O. Beckman

An investigation of conditioning in cats to multiple stimuli . . 211
By Charles Nelson Winslow, Robert Kantor, and Carl J. Warden

SHORT ARTICLES AND NOTES 227

A jumping method for establishing differential responses in pigeons (Ward Halstead and George Yacorzynski) . . 227

A note on the hiccough of the neonate (Isabelle F. Wagner) . 233

The sleeping posture of the neonate (Isabelle F. Wagner) . 235

A note on the perception of linear gestalten in the rat (I. Krechevsky) 241

APPARATUS 247

The block printing stick as a device for registering choice (Marie L. H. Forbes) 247

BOOKS 251

Lois Barclay Murphy, *Social behavior and child personality* (Florence L. Goodenough) 251

The Journal of Genetic Psychology, 1938, 52, 3-13.

A GENETIC STUDY OF THE BERNREUTER PERSONALITY INVENTORY*[1]

The Department of Psychology, Stanford University

PAUL R. FARNSWORTH[2]

PRELIMINARY REMARKS

As so much has been written about the Bernreuter Personality Inventory (1) the present writer can safely assume that little need here be said concerning its nature Suffice it to state that the inventory is a paper and pencil affair which at first purported to measure four personality variables: neurotic tendency (on its *B1-N* scale), self-sufficiency (on its *B2-S* scale), introversion (on its *B3-I* scale), and dominance (on its *B4-D* scale). Later, in the course of his researches, Bernreuter found that the *B1-N* and *B3-I* scales were measuring almost identical variables. Still later, Flanagan (3), by the technique of factor analysis, reduced the scales to two—a measure of confidence in oneself (*F1-C*), and a measure of sociability (*F2-S*).

Although a number of people have studied the inventory to learn of its validity and reliability, little or nothing has been done with it along genetic lines. And it was just this genetic angle that interested the present writer and motivated him to study the stability of responses over the college period In his research he also employed the R. C Peterson *Attitude Toward War* scale; the data derived from the giving of this attitude scale have been reported elsewhere (2).

---

*Received in the Editorial Office on March 30, 1937

[1] Reported in part at the 1934 meetings of the Western Psychological Association

[2] The writer gratefully acknowledges the financial aid of the Laura Spelman Foundation, and the kindness of Professors E E. Robinson, Reginald Bell, and M H Savelle, who allowed the writer to contact members of their classes in citizenship and history He also wishes to thank Drs. Quinn McNemar and Paul Buttenwieser for statistical advice, and the Messrs Leonard Ferguson, Gordon Williams and William Biel for their statistical and clerical assistance

PROCEDURE

In the autumn of 1932, 319 male students of Stanford University were contacted through the compulsory freshman course in citizenship. Fifty-five of these were retested the following autumn, 53 others in 1934, and 50 early in 1936, the contacting being accomplished by class solicitation or by letter. At the time the study began, women students were included. Shortly afterwards, however, there occurred a considerable change in the method by which Stanford's women students were selected. As it was not clear just what selective factors were operating, it was decided to confine the experiment to male subjects. To furnish a very rough check on general temporal (cultural) changes in score, 371 new-entering freshmen were tested in 1933, 83 in 1934, and 107 in 1935,[3] these scores were compared with those made by the experimental groups.

RELIABILITIES AND TOTAL SCORE RELATIONSHIPS

Odd-even reliabilities were figured for the six scales; they were found to resemble those reported by Bernreuter and by other researchers

TABLE 1

ODD-EVEN CORRELATIONS (STEPPED-UP BY BROWN-SPEARMAN)	
B1-N	.87±.01
B2-S	75± 02
B3-I	.79± 01
B4-D	.84±.01
F1-C	.80± 01
F2-S	76±.02

The intercorrelational figures (Table 2) show rather clearly that the *B1-N*, the *B3-I* and the *F1-C* scales measure very much the same variable, which is also tapped to a considerable degree by the *B4-D* scale (inversely, of course). The *B2-S* and the *F2-S* show considerable resemblance to each other.

The retest correlations can be seen in Table 3. It is obvious from the data of this table that the responses recorded in an inven-

---

[3] The last mentioned group came from a compulsory course in history which had superseded the course in citizenship

TABLE 2

INTERCORRELATIONS (NOT CORRECTED FOR ATTENUATION)

	B2-S	*B3-I*	*B4-D*	*F1-C*	*F2-S*
B1-N	— 38± 03	92± 01	— 72± 02	94± 01	.29± 04
B2-S		— 28± 04	.44± 03	—.55±.03	.67± 02
B3-I			— 58± 02	.88± 01	36± 03
B4-D				—.88± 01	.09± 04
F1-C					04± 04

TABLE 3

RETEST VALUES

Scale	1933	1934	1935-36
B1-N	76± 04	60± 06	.69±.05
B2-S	76± 04	74± 04	57± 07
B3-I	70± 05	.62± 06	.69± 05
B4-D	73± 04	73± 04	72± 05
F1-C	77± 04	.58± 06	66± 05
F2-S	74± 04	.56± 06	44± 08

tory such as the Bernreuter recur with considerable consistency over the college years. On the other hand, answers to an attitude questionnaire such as the Peterson War Scale mentioned above do not tend to remain unaltered. The latter's retest values were found by the present writer to range from .88± 02 (immediate retest) to 12±.09 (retest after somewhat over three years).

The formulae employed for finding the *PE*'s of the differences and deviations were: (*a*) where a portion of a group was compared with the entire group, the standard deviation of the entire distribution was divided by the square root of the smaller population, (*b*) where a group was compared with its retest, $\sqrt{PE^2_{m1} + PE^2_{m2} - 2r\ PE_{m1}\ PE_{m2}}$, (*c*) where no correlation existed between the groups, $\sqrt{PE^2_{m1} + PE^2_{m2}}$.

TABLE 4

DATA ON THE *Bl-N**

Means Compared	Differences and Deviations	*D/PE*
A vs *B*	+3.60± 3.87	.93
A vs. *C*	—29.56± 6.18	4.78
A vs *D*	+ .89± 5.85	.15
A vs *E*	.08± 4.56	.02
A vs *F*	.32± 4.81	.07
A vs *G*	4.29± 4.81	.89
A vs. *H*	7.22± 9.91	.73
A vs. *J*	12.46±10.11	1.23
A vs. *L*	20.69±10.41	1.99
E vs *H*	7.14± 9.88	.72
F vs *J*	12.78±10.07	1.27
G vs *L*	16.40±10.10	1.62
H vs. *I*	—15.82±13.36	1.18
J vs *K*	— 2.68±14.00	.19
L vs *M*	—32.20±13.48	2.39

*The "+" sign means that the change is in the direction of "more neurotic behavior" (according to the Bernreuter use of the term); "—", means "less neurotic." The figures without signs refer to samples and not to changes.

MEANS AND POPULATIONS

	A	*B*	*C*	*D*	*E*	*F*	*G*
Mean =	—76.69	—73.09	—106.25	—75.80	—76.77	—76.37	—72.40
N =	319	371	83	107	260	234	234
		H	*I*	*J*	*K*	*L*	*M*
Mean =		—83.91	—99.73	—89.15	—91.83	—56.00	—88.20
N =		55	55	53	53	50	50

There seemed to be no cultural change manifested in the 1933 data (*A* vs. *B*) or in those of 1935 (*A* vs. *D*), but a perceptible swing toward the "less neurotic" end of the scale in those of 1934 (*A* vs *C*). The phrase "less neurotic" merely means that the score was lower in 1934. However, as a number of researchers have determined (4) that the *average* score is *in reality* the least neurotic, this change, the only significant one found with the *Bl-N* scale, should probably be interpreted as toward *more* neurotic behavior. As reflected in the *Bl-N* scale there appeared to be no selective mortality at Stanford over the college period (*A* vs. *E*, *A* vs *F*, *A* vs *G*). The retest groups appeared to be fairly representative samples of the larger group tested in 1932 (*A* vs *H*, *A* vs. *J*, *A* vs. *L*). The retest groups appeared to be fairly representative samples of the larger groups still in school in 1933, 1934 and 1935 (*E* vs *H*, *F* vs *J*, and *G* vs. *L*). There were no significant changes in *B*1-*N* score over the college period (*H* vs. *I*, *J* vs *K*, *L* vs. *M*) This holds true even when the 1934 cultural shift is considered

TABLE 5

DATA ON THE *B*2-*S**

Means Compared	Differences and Deviations	*D/PE*
A vs *B*	—3.78± 2.52	1.50
A vs *C*	+10.95± 3.79	2.89
A vs *D*	—12.05± 3.76	3.20
A vs *E*	1.63± 3.01	.54
A vs *F*	3.09± 3.17	.97
A vs *G*	6.88± 3.17	2.17
A vs *H*	6.71± 6.54	1.03
A vs *J*	3.97± 6.66	.60
A vs *L*	6.65± 6.86	.97
E vs *H*	5.08± 6.39	.79
F vs *J*	.88± 6.17	.14
G vs *L*	.23± 6.51	.04
H vs *I*	+3.64± 7.21	.50
J vs. *K*	—5.85± 6.42	.91
L vs *M*	+7.80±10.05	.78

*The "+" sign means that the change is in the direction of "more self-sufficient behavior"; "—" means "less self-sufficient" The figures without signs refer to samples and not to changes

---

LEGEND FOR TABLES 4, 5, 6, 7, 8, AND 9

A refers to scores made by original freshmen group tested in 1932
B refers to scores made by freshman group tested in 1933 as a check on temporal change
C refers to scores made by freshman group tested in 1934 as a check on temporal change.
D refers to scores made by freshman group tested in 1936 as a check on temporal change
E refers to 1932 scores of those still in school in 1933
F refers to 1932 scores of those still in school in 1934
G refers to 1932 scores of those still in school in 1935-36
H refers to 1932 scores of those retested in 1933.
I refers to 1933 scores of those retested in 1933
J refers to 1932 scores of those retested in 1934
K refers to 1934 scores of those retested in 1934
L refers to 1932 scores of those retested in 1935-36
M refers to 1935-36 scores of those retested in 1935-36

MEANS AND POPULATIONS

	A	*B*	*C*	*D*	*E*	*F*	*G*
Mean =	34.25	30.47	45.20	22.20	32.62	31.16	27 37
N =	319	371	83	107	260	234	234
	H	*I*	*J*	*K*	*L*	*M*	
Mean =	27.54	31.18	30.28	21.43	27.60	35.40	
N =	55	55	53	53	50	50	

As none of the critical ratios was of the magnitude of 4 or over it is probable that the few changes present are "chance" affairs.

TABLE 6

DATA ON THE *B3-I**

Means Compared	Differences and Deviations	*D/PE*
A vs *B*	+3.19±2 38	1.34
A vs *C*	—20.79±3.82	5 44
A vs *D*	—4.71±1.62	1.30
A vs. *E*	.14±2.84	05
A vs *F*	.37±2.99	12
A vs *G*	.49±2 99	.16
A vs. *H*	4 37±6 16	.71
A vs *J*	3.74±6.28	.60
A vs. *L*	7.69±6.47	1.19
E vs. *H*	4.23±6.11	.69
F vs. *J*	4.11±6.18	.67
G vs. *L*	7.20±6 62	1 09
H vs. *I*	—6.54±6.94	94
J vs *K*	—6.23±7.66	81
L vs *M*	—14.00±7.78	1 80

*The "+" sign means that the change is in the direction of "more introverted behavior", "—" means "less introverted" The figures without signs refer to samples and not to changes.

MEANS AND POPULATIONS

	A	*B*	*C*	*D*	*E*	*F*	*G*
Mean =	—44.09	—40.90	—64.88	—48.80	—44.23	—43.72	—43.60
N =	319	371	83	107	260	234	234
	H	*I*	*J*	*K*	*L*	*M*	
Mean =	—48 46	—55.00	—47.83	—54.06	—36.40	—50.40	
N =	55	55	53	53	50	50	

Because of the high correlation existing between the scores on the *B1-N* and *B3-I* scales the data of Table 6 must closely resemble those of Table 4. The only significant change occurred when the 1934 freshmen were compared with those matriculating in 1932. The former were "less introverted" in Bernreuter's meaning of the term.

TABLE 7

DATA ON THE *B4-D**

Means Compared	Differences and Deviations	*D/PE*
A vs. *B*	+2.12±3.12	.68
A vs. *C*	+5.89±4.56	1.29
A vs. *D*	—18.63±4.42	4.21
A vs. *E*	.83±3.72	.22
A vs. *F*	.60±3.92	.15
A vs. *G*	3.63±3.92	.93
A vs. *H*	8.44±8.08	1.04
A vs. *J*	9.32±8.23	1.13
A vs. *L*	19.03±8.48	2.24
E vs. *H*	9.27±7.81	1.19
F vs. *J*	9.92±7.87	1.26
G vs. *L*	15.40±8.03	1.92
H vs. *I*	+14.19±8.29	1.71
J vs. *K*	+1.89±7.26	.26
L vs. *M*	+24.60±9.84	2.50

*The "+" sign means that the change is in the direction of "more dominant behavior", "—" means "less dominant". The figures without signs refer to samples and not to changes.

MEANS AND POPULATIONS

	A	*B*	*C*	*D*	*E*	*F*	*G*
Mean =	59.83	61.95	65.72	41.20	59.00	59.23	56.20
N =	319	371	83	107	260	234	234

	H	*I*	*J*	*K*	*L*	*M*
Mean =	68.27	82.46	69.15	71.04	40.80	65.40
N =	55	55	53	53	50	50

There appeared to be no cultural change in 1933 (*A* vs. *B*) or in 1934 (*A* vs. *C*) but an appreciable one in 1935 (*A* vs. *D*) in the direction of "less dominance" (in Bernreuter's use of the term). No other significant changes were visible unless it is considered that the 1935 cultural shift be included in interpreting the change in retest score from 1932 to 1935 (*L* vs. *M*). Quite possibly this latter change may be a meaningful one (toward "more dominant behavior").

TABLE 8

DATA ON THE *F*1-*C**

Means Compared	Differences and Deviations	*D*/*PI*
A vs. *B*	—8.18± 4.29	1.90
A vs. *C*	—40.84± 6.73	6.07
A vs. *D*	—2.58± 6.12	.42
A vs. *E*	1.86± 4.95	.38
A vs *F*	2.52± 5.22	.48
A vs. *G*	4.62± 5.22	.89
A vs. *H*	5.74±10.75	.53
A vs. *J*	9.21±10.96	.84
A vs. *L*	18.62±11.29	1.65
E vs *H*	7.60±10.61	.72
F vs. *J*	11.73±12.23	.96
G vs. *L*	14.00± 9.08	1.54
H vs. *I*	—20.44±10.85	1.88
J vs *K*	—2.08±13.25	.16
L vs. *M*	—27.60±14.20	1.94

*The "—" sign means that the change is in the direction of "more confident in oneself." The figures without signs refer to samples and not to changes

MEANS AND POPULATIONS

		A	*B*	*C*	*D*	*E*	*F*	*G*
Mean	=	—48.62	—56.80	—89.46	—51.20	—46.76	—46.10	—44.00
N	=	319	371	83	107	260	234	234
		H	*I*	*J*	*K*	*L*	*M*	
Mean	=	—54.36	—74.80	—57.83	—59.91	—30.00	—57.60	
N	=	55	55	53	53	50	50	

As the scores on *F*1-*C* correlate extremely well with those of the *B*1-*N* and the *B*3-*I*, the conclusions to be drawn from Table 8 must be practically identical with those from Tables 4 and 6. The only significant change occurred when the 1934 freshmen were compared with those matriculating in 1932. The former had "more confidence in themselves" (in Flanagan's use of the term).

TABLE 9

DATA ON THE *F2-S**

Means Compared	Differences and Deviations	*D/PE*
A vs *B*	+6.88 ±3.06	2.25
A vs. *C*	—6.72± 4.80	1.40
A vs *D*	+1.82± 4.37	.42
A vs *E*	1.98± 3.66	.54
A vs *F*	3.15± 3.86	82
A vs *G*	3.98± 3.86	1.03
A vs. *H*	9.44± 7.96	1.19
A vs *J*	9.21± 8.11	1.14
A vs. *L*	82± 8.35	10
E vs *H*	7.46± 7.90	94
F vs *J*	6.06± 8.58	.71
G vs *L*	4.80± 7.84	.61
H vs *I*	+6.25± 8.69	.72
J vs *K*	—10.37±10.74	97
L vs *M*	—1.20±14.10	09

*The "+" sign means that the change is in the direction of "less sociability"; "—" means "more sociability." The figures without signs refer to samples and not to changes

MEANS AND POPULATIONS

	A	*B*	*C*	*D*	*E*	*F*	*G*
Mean =	—22.02	—15.14	—28.74	—20.20	—24.00	—25.17	—26.00
N =	319	371	83	107	260	234	234

	H	*I*	*J*	*K*	*L*	*M*
Mean =	—31.46	—25.21	—31.23	—41.60	—21.20	—22.40
N =	55	55	53	53	50	50

As none of the critical ratios was of the magnitude of 4 or over it is probable that the few changes present are "chance" affairs.

ITEM CHANGES

It should be recalled that in this inventory there are three possible answers for each item. By "chance," therefore, 33.33 per cent of identical item answers can be expected over a period of time. This figure is to be contrasted with the following data: 71.35 per cent of the items were answered *identically after one year*, 65.45 per cent after two years, and 64.91 per cent after somewhat over three years.

As a sample of questions which were answered *differently* after a period of time might be mentioned "*Do you try to get your own way even if you have to fight for it?*" Here the major changes were from "*Yes*" to "*?*" in 1934, and from "*No*" to "*Yes*" in 1935. As a sample of questions which tended to be answered *identically* after a period of time might be mentioned "*Do you dislike finding your way about in strange places?*"

Résumé

In an attempt to observe long-time changes in answers to questions of the adjustment inventory sort, male freshman students at Stanford University were tested in 1932 on the Bernreuter *Personality Inventory*, and were then subdivided for later retesting. The subsequent retesting took place at intervals of one, two and somewhat over three years.

Additional freshmen were tested in 1933, 1934 and 1935 This procedure was deemed necessary to guard against the possibility of cultural changes of a general sort. Slight shifts were found to be present on the *B1-N* scale in 1934 (toward the "less neurotic" end of the scale according to Bernreuter's theory but toward the "more neurotic" according to the present writer's), on the *B3-I* in 1934 (toward the "more extroverted" end), on the *B4-D* in 1935 (toward the "less dominant" end), and on the *F1-C* in 1934 (toward the "more self-confident" end).

Odd-even reliability and intercorrelational values were found which closely resembled those reported by other researchers.

No significant changes occurred during the years at Stanford on any of the six scales.

The one year retest values ranged from .70±.05 to .77±.04 (median approximately .75); those for two years, from .56±.06 to 74±.04 (median approximately .61); those for somewhat over three years, from .44±.08 to .72±.05 (median approximately .67). These values are *far* higher than were those previously reported for attitude scales.

Also, 71.35 per cent of the items were answered in an identical manner after one year had elapsed, 65.45 per cent after two years, and 64.91 per cent after somewhat over three years. This type of pencil and paper test yields symbolic responses which are far more stable (unchangeable) than are those of the "attitude toward war" variety

REFERENCES

1. BERNREUTER, R. G. The personality inventory. Stanford University: Stanford Univ. Press, 1931.
2. FARNSWORTH, P. R. Change in "attitude toward war" during the college years. *J. Soc. Psychol.*, 1937, 8, 274-279.
3. FLANAGAN, J. C. Factor analysis in the study of personality. Stanford University: Stanford Univ. Press, 1935.
4. LAPIERE, R. T., & FARNSWORTH, P. R. Social psychology. New York: McGraw-Hill, 1936. Pp. 357.

Department of Psychology
Stanford University
California

The Journal of Genetic Psychology, 1938, 52, 15-64.

SCORING THE RORSCHACH INK-BLOT TEST*

Development Health Inquiry, Western Reserve University

MARGUERITE R. HERTZ

While the Rorschach ink-blot test has interested the specialist engaged in the field of personality since its inception in 1921, it has not progressed as rapidly as had been anticipated. Several reasons for the retardation in the development and application of the test have been suggested by various investigators from time to time It is claimed, for example, that there is no adequate guide or manual to direct the examiner The method of administration and of scoring has not been sufficiently standardized Norms are grossly inadequate. The various claims for many of the diagnostic categories have not as yet been reliably and statistically confirmed

The Rorschach Test was originally based on an experiment with 405 subjects, 231 men and 174 women, of whom 117 were "normal persons" and the others mentally disordered cases. In the original manual (25), the method is explained and the early experimental results summarized. The method was subsequently described and elaborated by Behn-Eschenburg (4), Löpfe (17), Müller (21), Loosli-Usteri (16), Pfahler (24), Juarros and Soriano (13), Oeser (23), Beck (1), Vernon (30), Salas (28), Mandowsky (18), Kerr (14) and Monnier (20).[1]

The method of administration of the Rorschach test was treated in another paper by the writer[2] It was pointed out that the administration has not been standardized to the point where all examiners apply the method in the same way Variations in instructions, in the method of presenting the blots, in the accompanying conversation and suggestions, in the attitudes of the subject, in the length

---

*Received in the Editorial Office on April 3, 1937.

[1] For a summary of the work done up to January, 1934, reference should be made to Vernon, P E, *Brit J. Med Psychol*, 1933, 13, 89-118, 179-200; 271-291, or to Hertz, M R., *Psychol Bull*, 1935, 32, 33-66 For references from then to the present time, see Vernon, P E, *J Ment Sci*, 1935, 81, 894-920

[2] Hertz, M R The Rorschach ink-blot test. Ch. IV, unpublished thesis, 1931 Western Reserve University Library, Cleveland, Ohio. To appear in *Child Development*

of time allotted—all affect the resultant psychogram. There can be little doubt that the test can progress only if a standardized method be developed which may be used universally by all Rorschach examiners.

In similar manner, the method of scoring has not been developed adequately. Rorschach responses are scored according to four categories: (*a*) the mode of apperception (*W*, *D*, *Dr*, *Do*, *DS*), (*b*) the quality of the apperception (*F*, *M*, *C*, *CF*, *FC*, *F* [*c*]), (*c*)

TABLE 1

KEY TO THE RORSCHACH SYMBOLS

Symbols German	 French	 English	Interpretation
I. *Mode of apperception*			
G	G	W	whole apperception
dG		dW	whole apperception determined from interpretation of detail
D	D	D	normal detail, i.e., part of blot which is frequently perceived
Dd	Dd	Dr	rare detail, i.e., part which is rarely or unusually perceived
Do	Do	Do	oligophrenic detail, i.e., interpretation of a part of a body where the normal subject gives the whole body
Dzw	Dbl	DS	space detail
—	—	DrS	rare space detail, not used by Rorschach
Sukz	Succ	Succ.	order in which the different modes of apperception follow each other
Erft	Types de Perc.	App.	formula expressing the relative number of different modes of apperception
II. *Quality of apperception*			
F	F	F	form apperception—stimulated by form of blot
F+	F+	F+	sharply perceived forms
F—	F—	F—	poorly perceived forms
B	K	M	movement, kinesthetic interpretations
Bkl.	Kp.	Md	interpretation of detail involving movement
F(Fb)	F(C)	F(C)	chiaroscuro—interpretation determined by differential shading of black-gray, more rarely color
Fb	C	C val 1½	interpretation determined by color only
FbF	CF	CF val. 1	color-form, interpretation determined by color first and then form
FFb	FC	FC val. ½	form-color, interpretation determined by form first and then color

TABLE 1 (*continued*)
KEY TO THE RORSCHACH SYMBOLS

Symbols German	 French	 English	Interpretation
ΣFb	ΣC	ΣC	total color score, according to the numerical values assigned
Farbenschock	Choc-couleur	Color shock	Hesitation or stupor at appearance of first color picture, more rarely with other colored pictures
%F+	%F+	%F+	$\frac{100\ F+}{(F+)\ +\ (F-)}$
Erlebnistypus	Type de caractere	Type of Psychic (Affective) Reaction	Formula expressing the relations between movement and color scores
$xB > yFb$	$xK > yC$	$xM > yC$	Introversive type, movement being greater
$xB < yFb$	$xK < yC$	$xM < yC$	Extratensive type, color being greater
$xB = xFb$	$xK = xC$	$xM = xC$	Ambiequal type, color being equal to movement
OB, OFb	OK OC	OM OC	Constricted type, no color or movement
xB, yFb	xK yC	xM yC	Dilated type, many color and movement
(x and y represent different numerical quantities)			
III *Content of apperception*			
T	A	A	entire animal form
Td	Ad	Ad	part of an animal form, not an anatomical reference
%T	%A	%A	percentage of stereotypy, $\frac{100\ (A+Ad)}{R^*}$
M	H	H	entire human form
Md	Hd	Hd	part of human figure, not anatomical
%M	%H	%H	percentage of human figures $\frac{100\ (H+Hd)}{R^*}$
Anat	Anat	Anat	strictly anatomical forms
(corresponding translations)		abstr	abstractions
		alph	alphabetical forms, letters
		bot	botanical forms, plant life, flowers, trees
		cave	cave forms, holes, valleys
		color	colors
		cl	clouds, mist, fog
		fire	fire forms, explosions
		ft	fountains
		geog	geographical forms, maps (also under Wr.)

TABLE I (*continued*)
KEY TO THE RORSCHACH SYMBOLS

Symbols German	French	English	Interpretation
		geol.	geological forms, rocks
		ice	ice forms
		mt	mountains, hills
		no.	*numbers*
		obj	small concrete objects, not otherwise classified
		Sc	landscape, scenery, generalized scenes
		st	statues, monuments
		str.	structures, architecture
		sym.	symbolic *forms*
		vol.	volcanos
		wr.	water forms, rivers, lakes, seas
		( )	parentheses indicate imaginary forms
		etc.	
IV. *Originality of apperception*			
Orig.	0	0	Original answers
Orig +	0+	0+	good original answers
Orig —	0—	0—	poor original answers
%Orig.	%0	%0	percentage of original answers $\frac{\text{Original}}{R^*}$
I	I	I	individual answers
V	Banal	P	Popular or common answers
Antw.	R	R*	Total number of content responses

the content, and (*d*) originality (*O*) or banality of response (*P*). The Rorschach symbols, abbreviations and formulae herein used are those employed at the Brush Foundation. Table 1 summarizes them, giving their equivalents in German and French studies.

In order to score, it is obvious that norms or standards are necessary on which to base judgment. The *F*+ for example, according to Rorschach is the form given by the majority of normal subjects with the greatest frequency The normal detail (*D*) is that which is selected most frequently by the majority of people while the original answer (*O*) is the response given by one person in one hundred Obviously, the norms for these factors must be ascertained before any scoring can be done.

Unfortunately, such norms are not available in published form in sufficient number. The few standards that are to be found have been based on small and highly selective groups and have not as yet been universally accepted. It is indeed curious that despite this deplorable lack of norms for scoring the test, it has been extensively used both in research and clinical practise.

The scoring procedure generally followed by most examiners is to refer to the original manual (25) and to Oberholzer (26) and to score responses by comparison with the samples and suggestions offered in these publications. It should be pointed out that the manual contains only 27 records Scoring in this manner is highly indefinite and uncertain, different criteria being applied by different examiners who must rely upon their experience and "intuition" Oberholzer (26) says, for example, "formulations cannot be rigidly classified Even with great experience, the subjective conclusion by analogy cannot be entirely avoided."

It is obvious that no matter how conscientious the examiner may be, inaccuracies are bound to occur. Many examiners try to reduce the amount of error by helping each other define criteria for scoring. Thus Beck and Levy (1) checked on each other, and Oberholzer checked on them. Meltzer (19) reports that he scored his records for stutterers in the light of Rorschach literature supplemented by suggestions and advice from Beck. Other examiners refer to the samples of responses which are included in various Rorschach publications.

Obviously, this method of procedure is loose As Beck (3) observed abroad, Rorschach examiners took "liberties which are inconsistent with and foreign to the scientific approach." Practical application to clinical procedure must be seriously hampered by such lack of uniformity. Dimmick (33) for example recently tried to apply the test to three clinical types of dementia praecox and concluded that the test will be valuable for personality diagnosis only when more objective classification of the responses is possible. In working with the test, he found the scoring too subjective, especially for *F*+, *F*—, *Do*, *O*+ and *O*—

Some examiners use lists of norms which they have amassed. Such lists have been published by Loosli-Usteri (16), Linares Maza (15), Salas (29) and Gardner (8). Meltzer (19) also reports that he adhered to statistically prepared tables which have not, however, been published. In the most recent summary on the Rorschach test, Vernon (31), still recognizes the need for tables of responses from representative groups in order that some degree of objectivity in scoring be attained.

There are some Rorschach examiners, however, who prefer the subjective, artistic, and intuitive approach. Wells (32), for example, feels that despite the subjectivity of the Rorschach method, it is one which should be emphasized "if the theme of 'personality' tests is to get out of the tangle in which the premature recourse to statistics (or in less amiable terms, the substitution of statistics for insight) has enmeshed it." He points out, therefore, that the administrator and the interpreter of the test must be artistic and intuitive and must display a delicate critical sense.

One generally experiences difficulty in trying to get an objective score in any free association test. The test is purposely designed to permit wide latitude of response and free play to the imagination The ink-blot test is no exception. Subjective evaluation is inevitable and even desirable in interpreting the final score. But a test to be reliable must exhibit some degree of uniformity and must be amenable to a fairly objective scoring so that it may be used efficiently by all test administrators Therefore, while no attempt to adhere to a rigid stereotyped procedure should be made, it should be regulated to a degree of efficiency Further, although subjective interpretation of the test results is desired ultimately, an objective and adequate scoring will reduce the necessity for subjective evaluation and thus place the test within the bounds of reliability and practicality.

Many investigators as Müller (21), Löpfe (17), Beck (1-3), and Vernon (30, 31) have recognized these serious limitations of the scoring procedure and have tried to outline criteria for scoring certain of the Rorschach categories. Hertz' summary (11) of the literature up to 1934 indicates the gross inadequacy of the criteria which have been expounded to that date. Meltzer (19) has subsequently suggested certain improvements and modifications in the scoring system which would tend to increase the objectivity and reliability of the scoring As already indicated, Vernon's latest resumé (31) emphasizes the uncertainty and subjectivity of the Rorschach scoring even to this date

Since it was recognized that one of the most urgent needs for the Rorschach test is the standardization of the method of administration and scoring, especially the determination of objective criteria for scoring and the development of adequate tables of responses from a representative group, the Brush Foundation[8] undertook such standardization as part of its research project on the basis of the Rorschach records of 300 Junior High School students, 150 boys and 150 girls, selected as representative of average American-born white children. These subjects were selected at random as to chronological age, class standing and school grades. Reference should be made to a previous paper for a description of the group which shows distributions according to chronological age, sex and intelligence quotients, all of which approximate the normal curve (12)

It is obvious that the success of any scoring method devised depends in large measure on the facility, accuracy, and the uniformity of the administration of the test It was therefore necessary to perfect if possible the method of giving the test before attacking the problem of scoring.

Reference has already been made to the report on the standardization of the method of administration as developed at the Brush Foundation. In brief, Personal Data Sheets, Record Blanks, and Summary Sheets were prepared for quick and efficient recording of all necessary data. Systematic and uniform directions were prepared and memorized. A trial blot was introduced to make the responses to the first card more comparable with the rest of the responses, and to establish a more favorable mental set at the beginning of the test.

Diagrams were made for each blot, parts of the blot being assigned symbols in the form of letters and numbers for indicating the exact location of the detail selected for interpretation. These diagrams were utilized in giving the test. Figure 1 is a reproduction of the diagram for Card II. It may be observed that (*S*) represents the side figures, (*3*) the upper red figures, (*4*) the bottom middle red, (*2*) the middle space form, (*12*) the black projection top center, etc. In addition the "V" was used to show the position in which the card was held, the apex always representing the top of the card Answers were recorded, using these symbols, thus:

Card II "*two men dancing (W)*"
"*a clown (S) with a red hat (3)*"
V "*a king's head (7)*"
"*a top (2)*"
V "*a peacock (25)*"
"*a rabbit (5)*"

---

[8] The Brush Foundation is the central partner in the program known as the Developmental Health Inquiry of the Associated Foundations, Western Reserve University.

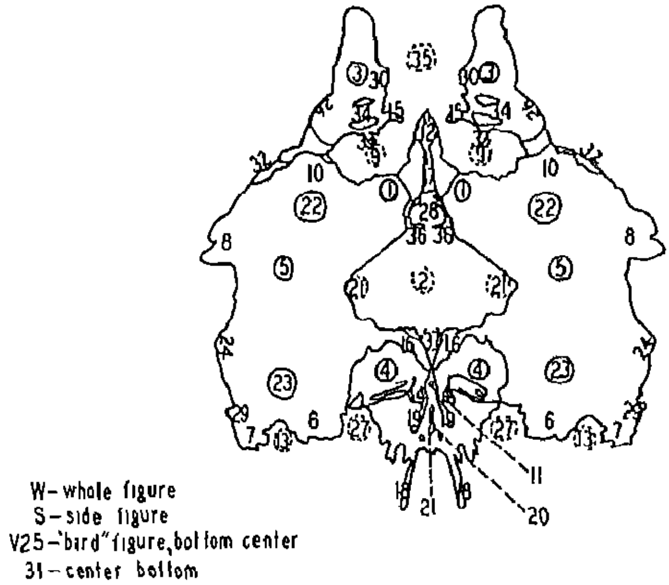


FIGURE 1
Diagram for Card II

Finally, the reaction time was recorded from the time the card was presented until the first response was given. Two minutes were allotted for each card.[4]

Problem. Standardization of the Method of Scoring

The next problem was to determine upon a standardized method of scoring. The responses of the 300 students were therefore subjected to statistical analysis and qualitative study, and a method of scoring was evolved which would enable different examiners to score Rorschach records in the same way.

Frequency Tables for each blot were constructed containing all the responses. Construction of these tables involved:—

I. Preparation of Catalogues in which answers were recorded on separate cards and arranged according to their respective location on the blots,

---

[4] For further details on the method of administration of the Rorschach Test as modified at the Brush Foundation, refer to note 2.

II. Preparation of "Preliminary Lists" on which answers were so arranged that their numerical frequencies might be readily computed,

III. Determination of the criteria for the evaluation of the test factors,

IV. Final determination and construction of the Frequency Tables for use in scoring.

I. Cataloguing Responses

Each subject had been assigned a serial number, 1 through 300. Every response given by a subject was recorded on a separate catalogue card by using this serial number. The location of the portion of the blot selected for interpretation was indicated in the upper left-hand corner by means of the symbol used on the diagrams. For example, for Card II, a response of *"a sunset over the mountains near a lake"* was recorded in its entirety on a card labelled in the upper left-hand corner "V II W" meaning card held upside down, Whole.

Not only was every response recorded in this manner, but every *item* of a response, an item being defined as every *form* which had actually been given by a subject. For example, the above response was also recorded on separate cards thus: "V II 4", "V II 5" and "V II 2".[5] For this one response, then, four cards were prepared. In this manner, every form and every response which had been expressed verbally appeared on separate cards, the location appropriately designated in the corner. Below each answer or item of an answer, there appeared in columns the serial numbers of the subjects who had given that reply. If a particular answer happened to be the first reply given, the serial number was written to the left in the red ink; if it was not the first answer, it was written to the right in green ink.[6] All serial numbers of items were enclosed in parentheses to indicate that these did not constitute an entire answer. Table 2 is a copy of two index cards with their respective tabulations taken from the catalogue files.

---

[5] The symbols used refer to those adopted by the Brush Foundation and are taken directly from the prepared diagrams. See Figure 1.

[6] Answers were thus classified on the basis of whether they were given first or subsequently. Throughout this paper they are accordingly designated as "first answers" and "other answers." This was done in order to study the test from the viewpoint of the first reaction given to the blots. This subject matter has been deferred for treatment upon another occasion.

TABLE 2

Reproduction of Two Index Cards Used in the Tabulations for the Frequency Tables

V W Butterfly

* 4	40	73	111	139	163	191	222	249	286	** 1	140
9	41	74	114	142	164	194	226	250	283	24	172
12	42	78	115	144	165	195	232	256	289	26	178
13	47	84	116	146	167	196	233	257	290	43	187
14	48	89	117	147	168	198	234	258	292	46	200
16	50	91	119	148	171	199	235	259	294	55	201
18	51	95	120	150	173	201	236	260	298	57	208
19	55	98	123	152	174	202	237	263	299	61	225
20	58	99	125	153	176	207	240	264		67	253
23	60	103	126	157	177	211	241	270		90	255
25	62	104	130	158	179	217	242	271		102	279
29	63	107	131	159	181	218	244	275		106	280
24	66	108	134	161	185	219	245	278		109	288
37	69	110	138	162	188	221	248	285		122	

X (20) Man with hands up, over head

* 1	**24	** 114
(42)	36	198
167	42	
173	59	205
		207
(175)	(108)	(212)
188		214
191		241
195		252
209		(261)
300		265
		275
		290
		293

* Numbers appear in red, indicating first answers
** Numbers appear in green, indicating "other" answers

Yellow index cards were used as indicators to separate the respective parts of the blot which had elicited a response. Each yellow card was labelled on the projecting flap with the symbol of the detail to which it referred, corresponding to that on the prepared diagrams. Thus for each ink-blot, *location-divisions* were prepared, separated by yellow indicators. A section for whole answers always preceded those for details.

The catalogue cards with answers were placed in their proper location-divisions. Within each location-division, answers were classified into *content-groups* according to the content of the responses,—human, animal, anatomical, natural and artificial forms, these divisions being separated by colored markers, red, blue, green, black and yellow respectively. Within these content-divisions, answers were arranged alphabetically.

Cards were next regrouped in each content-division, according to similarity of forms ("form-class") following the scheme—

- I. Animate forms
  - A. Human forms, whole or parts
  - B. Animal forms not human including
    - 1. four-legged animals
    - 2. bats or birds
    - 3. crustaceans as crabs or lobsters
    - 4. small fish
    - 5. large fish (whales, seals, etc., different in form from no. 4)
    - 6. frogs and toads
    - 7. turtles
    - 8. insects or bugs (flies, beetles, butterflies, moths)
    - 9. snake forms (reptiles, crocodiles, lizards, alligators, snakes)
    - 10. worm forms (worms, caterpillars)
    - 11. sponges
    - 12. protozoa and unicellar animals
- II. Anatomical forms
- III. Inanimate forms
  - A. Single items
    - 1. natural forms (such as botanical forms, clouds, geographical forms, mountains, water forms, etc.)
    - 2. artificial forms
  - B. Generalized scenes

All cards were arranged according to these new divisions in the content-division, in alphabetical order.

II PRELIMINARY LISTS

After all answers and items of answers had been recorded in their appropriate categories, lists were prepared for each test blot with 15 columns providing for the following notations.

- I. Location symbol
- II. Position symbol
- III. Response
- IV. Reference numbers (in blue) to notes
- V. First Response, number of times given as a first answer
- VI. First Response, number of times given as an item incorporated in a first response
- VII. Other Response, number of times given as an entire answer, not the first

TABLE 3

Preliminary List Card V

Symbol Loc. I	Pos II	Response III	IV	1st R V	VI	Other R VII	VIII	Sum IX	X	XI	XII	Score XIII	XIV	XV
		HUMAN FORMS												
W		one person**	1*	1		10		11		11	W	F+	H	**
		dancer in costume		0		1		3						
		dancer with wings		0		2								
		devil with horns and wings		1		0		1						
		dwarf with wings		0		1								
		dwarf holding slate together		0		1		2						
		girl dressed up		0		1		3						
		girl with flowing gown		0		1					W	F+	(H)	O+
		girl with dancing veil		0		1								
		person with robes		0		1		2						
		person with long coat		0		1								
W		two persons**	2*	0		6		6		6	W	F—	H	**
		double angel		0		1		1						
		clowns coming together		0		1		1						
		men fighting		0		1		1						
		men with horns and dresses		0		1		1						
		persons meeting		0		1		1						
		persons diving												
		persons in fairy tales		0		1		1						
W	V	two persons, card reversed**		0		2		2		2	W	F—		**
		dwarfs lying down, feet propped against a pillar		0		1		1		1	W	F—	(H)	O—
		dwarfs lying down, feet up		0		1		1		1	W	F—	(H)	O—

TABLE 3 (*continued*)

PRELIMINARY LIST CARD V

Symbol				1st R		Other R		Sum				Score		
Loc I	Pos II	Response III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV
		ANIMAL FORMS												
W		one animal seen**	3*	0		5		5		5	W	F—	A	**
		animal		0		2		2		2				
		animal over fireplace, stretched out		0		1		1		2				
		dog stretched out		0		1		1						
		rabbit		0		1		1		1				
W	AV	animal skin**		0		2		2		2	W	F—	Ad	O—
		animal skin		0		1		1		2				
		muskrat skin		0		1		1						
W		two animals**	4*	2		4		6		6	W	F—	A	**
		animals clashing into each other		0		1		1		3	W	F—	A	O—
		animals charging toward each other		1		0		1			W	F—	A	O—
		animals fighting		1		0		1			W	F—	A	O—
		cows, heads cut off, put together, stretched out		0		1		1						
		dogs, feet out		0		1		1						
		goats charging together with their heads		0		1		1			W	F—	A	O—
W		flying animals**		0		3		3		3	W	F+	A	
		flying foxes		0		1		1		1				
		flying rats		0		1		1		1				
		flying kangaroos with fake wings		0		1		1		1				

* reference numbers appear in blue

** frequencies for the form-classes appear in red

VIII Other Response number of times given as an item incorporated in other responses
IX Sum of all entire answers given (V + VII)
X Sum of all items given (VI + VIII)
XI Total of all items and answers (IX + X)
XII. Score for W, D, Dr, Do, DS
XIII Score for F+ and F—
XIV Content symbol or abbreviation
XV Score for O+, O—, P.

Table 3 reproduces one page of these preliminary lists taken from the records of Card V. Responses were written in appropriate columns It will be observed that all the answers pertaining to "a person" were listed in alphabetical order under that heading which represents the *form-class*. The frequency for each response and for the form-class was computed, the latter being written in red ink. For example, an answer in terms of "a person" was given once as a "first answer," ten times as "other answers," never as "items," the total being eleven times in all. Similarly, "flying animals" was given no times as "first answer," three times as "other answers," totalling three times in all.

It should be noted that Column IX gives the total number of answers given to the blot and was the figure used in determining the distinction between normal and rare details. Column XI represents the number of times any reference was made to a particular kind of response and summarizes all the forms given. It was this figure which was used in determining good or poor forms and originality or commonness of response.

In this manner, lists of all responses were arranged for each card On the basis of these, definite criteria were formulated for the different test factors

III DETERMINATION OF CRITERIA FOR THE EVALUATION OF THE RORSCHACH TEST FACTORS

A *Scoring the mode of apperception.*

1. *Whole answers (W).*

Answers were scored "whole" where the subject indicated that the entire blot had been interpreted, not a part of it. Frequently a whole was given at once, thus: *"a bat," "a flower," "an angel,"* no details being mentioned. At other times, the whole figure was interpreted followed by descriptive details, as for example for Card

III, *"two butlers serving, holding tray, bodies bent, thin necks, and funny noses"* At times, details were seen successively and then all grouped into a whole coordinated idea, as for Card I, *"a witch at the side, another on this side, a pole in the center—why it must be witches burning something at the stake."* These answers were scored *"W"* When, however, it was evident that the whole had been built up from a detail or a number of details first, the answer was scored *"dW"* meaning built up from one or more details.[7]

2. *Detail answers, normal (D) and rare detail (Dr).*

All answers to parts of the blots were scored detail answers Distinction had to be made between the normal details (*D*) and the rare details (*Dr*)

Rorschach's instructions were "normal details are those selected most frequently by a normal group of subjects" (27). This suggests a quantitative and statistical criterion Beck (3) points out that there are two possible criteria for the *Dr*, "*a*) unusualness, such as is inherent in size, position, or a qualitative feature, *b*) infrequency." The qualitative criterion has not as yet been experimentally investigated.[8] Most workers use the quantitative standard

Lopfe (17) called all details normal, interpretations of which made up 1/22 of all those given to the blot Those details, the frequency of which was less than 1/22 of the total were considered rare details For Card II, for example, he received 465 responses from his subjects He took as his dividing line between normal and rare details, 21 or 1/22 of 465 He decided that all parts of the blots which were selected for interpretation more than 21 times were normal He explained that he selected this fraction because it gave results which were most like those which had been obtained by Rorschach with adults and he believed it would be practical to have similar results for comparative purposes

Lopfe's criterion of 1/22 was accepted by Loosli-Usteri (16) in distinguishing her normal and rare details. She had to use his criterion since her study involved a comparison of her data with his.

---

[7] Beck (2) has shown that some blots tend to elicit more Ws than others It is therefore necessary to weight this factor according to the blot. He lists values statistically computed to be used instead of W.

[8] In the first issue of the "Rorschach Research Exchange Service" (35), Dr Klopfer and his associates discuss qualitative criteria for this category

Even using his criterion, however, her normal details did not agree entirely with those of Rorschach or with those of Lópfe.

Beck (3) claims to use a quantitative criterion, also, though he thinks it necessary to investigate experimentally which is the more valid, the quantitative or qualitative. Vernon (30) finds this criterion of frequency unnecessary. He suggests another scheme of scoring, namely *W*+ to refer to the whole blot, *W*— to the greater part of it, *D*+ to a large portion and *D*— to a small detail, the examiner computing under which of these categories the response should go.

Gardner (8) decided to score as rare (*Dr*) all portions of the blots mentioned less than five times by his 300 subjects.

The Brush investigation adopted a quantitative criterion. The actual number of times specific details were selected for interpretation were tabulated for each blot, frequency distributions prepared, graphs plotted and studied to locate where the break seemed to come between the most frequently selected details and the rarely selected details. Figure 2 shows the graph for Card II, Table 4 the summary sheet upon which it is based.

TABLE 4

SUMMARY SHEET: FREQUENCY WITH WHICH DETAILS ARE SELECTED FOR INTERPRETATION

Card II. Symbol of Detail		Total no. of times selected	Score based on 20 percentile criterion
8	side figure	88	D
3	red figure on top	136	D
30	"heel" of red figure	2	Dr
32	narrow red just above black	3	Dr
15	"leg" of red figure, top	5	Dr
15+33	"legs" of red figure, together, top	5	Dr
33	"leg" of red figure, top	5	Dr
5	black figure at sides	237	D
4	bottom red figure	139	D
2	middle space form	75	DS
5+4	black plus bottom red part	2	Dr
8	projection, outer edge of black	30	Dr
24	profile, outer edge of black	1	Dr
22	upper half of black figure	6	Dr
10	"head" part of black figure	3	Dr
1	"paw" of animal toward center	28	Dr
1&16	"two paws" pointing in middle	11	Dr
12	projection, center top	66	D
12+2	projection top center with space form middle	2	D+DS
17	narrow lighter gray streak in top projection center	2	Dr

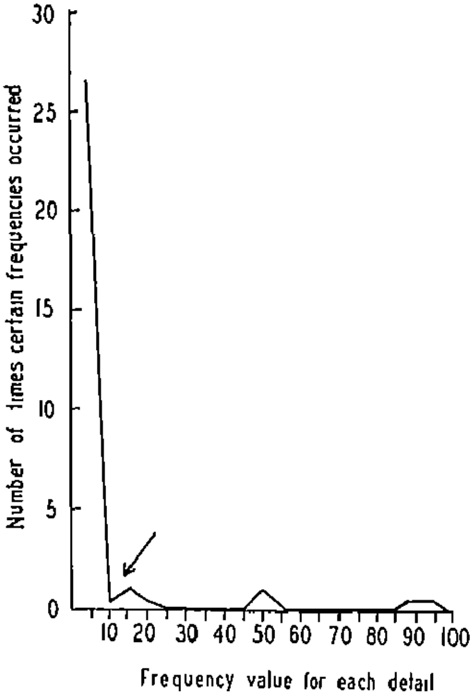


FIGURE 2
CARD II — FREQUENCIES WITH WHICH DETAILS ARE SELECTED FOR INTERPRETATION

Various percentiles were located for each distribution The 20th percentile was selected as criterion, since it appeared most consistently to be near the breaking point of the line. All details with a frequency above the twentieth percentile were considered normal details Thus in Card II, Table 4 shows that the Side Figure (*S*) was selected for interpretation 88 times, the red figure top-(*3*), 136 times, the black figure-(*5*) sides, 237 times, the bottom red figures, 139 times, and the projection center top-(*12*), 66 times These were considered normal details, the others, rare details

Table 5 presents the list of normal details for each ink-blot as determined by the criterion adopted.

TABLE 5

SUMMARY OF NORMAL DETAILS AS DETERMINED BY THE 20TH PERCENTILE CRITERION

Card		Location*
I	S	Side figure
	C	Center figure between sides
	21	Upper third of side figure
	2	Projection of side figure, "wing"
	31	Upper third of center figure
	15	Lower two-thirds of center figure, "bell"
	17	Darker form in center "lady's figure"
II	S	Side figure
	5	Black figures at sides
	3	Red top figure
	4	Bottom red figure
	12	Black pointed projection, middle top
	2	Middle space form
III.	S	Entire side figure
	5	Red middle, "butterfly"
	4	Red figures, upper corners
	1	Black part, middle bottom, "negro heads" card reversed
	2	Side figure, "leg"
	3	Middle part of black side figure
IV	1	"foot" figure, sides
	2	Middle projection
	4	"Snake" projections, sides
	3	Top part of "foot" figure, "toe" part
	5+5	top center, "face"
V.	S	Side figures
	2	Two bottom projections, middle
	1	Top middle figure, "head of rabbit"
	13	Side figures without extremities
(VI continued)		
	5	"Head" part of side figure, card reversed
	24	Lighter gray figure, center, bottom
	22	Top of upper third, "snake's head"
	2	Large projection extending outward at sides
VII	1	Top third of side figure
	2	Middle third of side figure
	3	Bottom third of side figure
	1+2	Top two-thirds of side figure
	3+3	Bottom third of whole figure
VIII	1	"Animal" figures at sides
	4	Top middle gray
	3	Blue middle
	20	"Rib-figure" or "fish bone," middle of gray blue
	2	Reddish top of red-orange, Middle bottom
	2+5	Red-orange, middle bottom
IX.	1	Brown part, top third
	2	Green part, middle third
	3	Pink part, bottom third
	10	"animal's face" in green on border of brown
	cl	"claws" in brown, projections extending into center
	13	Pink bottom, outer portion
	3+C+3	Pink part, bottom with center figure, "tree," card reversed
X	1	Pink portions
	9	Gray figure, center top
	9+9	Gray figures together at top, including projection (30)
	5	Blue figures, center

TABLE 5 (*continued*)

Card	Location*	Card	Location*
6	One of side extremities "leg" figure	3	Green figures, center bottom
30	Profile, under contour	12	"Rabbit's face" in green center bottom
VI S	Side figure, half of lower two-thirds of whole	20	Lighter green, center of green figure bottom, when reversed, "a person"
S+S	Lower two-thirds of whole		
1	Upper third of figure	6	Blue figures, sides, "spiders"
C	Center part through	4	Gray figures, sides
6	Darker black part of center top, "snake" figure	8	Orange figures, sides, bottom
		7	Yellow figures, middle

*Location symbols refer to diagrams used at the Brush Foundation for recording and locating responses. See Figure 1. These diagrams have been reproduced in Booklet containing the Frequency Tables to be used in Scoring the Rorschach Ink-blot Test which can be obtained from the Brush Foundation, Cleveland, Ohio.

3. *Space details* (*DS*)

The space details gave little difficulty. It was generally clear when a space form was selected for interpretation. The subject either pointed to the space form giving his answer or indicated the space upon cautious questioning of the examiner. For example, Card II, middle space form, *"a top,"* Card VII, middle space form, *"a lamp,"* Card I, white around the black silhouette, *"the ocean,"* all were readily recognized as space forms. When there was any doubt as to whether space details were involved, the following question was applied to the response: *"Could that answer have been given if the space forms had not been specifically seen?"*

When it was evident that the whole blot had been interpreted in conjunction with the space forms, symbol *"W(DS)"* was used. Oberholzer and Baenziger (26) used the symbol *"Gzw"* which corresponds to this. When space forms seemed to be more dominant, *"DS(W)"* was employed. When the white spaces seemed to be entirely responsible for the answers as in *"ocean," "water,"* or *"lake,"* they were designated as *"DS."*

4. *Oligophrenic details* (*Do*).

The oligophrenic details (those details where only parts of bodies are given where most normal subjects see whole bodies) were easily scored, once the lists were tabulated. Since wholes and parts of wholes were listed and placed together, it was readily determined from the numerical frequencies whether whole figures had been given more frequently than parts. In Card I, for instance, "a wing" for the side projection was scored *Do* because most subjects saw the entire figure, either as a *"person with wings,"* or *"a bird with wings,"* or an *"angel with wings."* Consulting the lists, the numerical frequency for *"wing"* was so much less than that for the other whole figures, that it was evident that it was the rare answer rather than the usual. Similarly, for Card V, *"a wing of a butterfly"* for the side projection was a *Do* because the numerical frequency for the whole butterfly showed that most people see the entire form rather that just a part. Again, in Card VIII, *"feet of some animal"* for the *"feet"* of the side figure which is generally seen as a whole animal was scored *Do.*

Many responses of parts of bodies were not oligophrenic details in Rorschach's sense, because it was more "normal" to give these parts than the wholes. Again, the numerical frequencies on the lists showed this. Card VII, top third, for example, was commonly interpreted as *"a person's head,"* few subjects giving a *"whole person"*" for that side of the blot. Hence this was not considered a *Do* response but rather a *D* response.

5. *Summary*

Once these criteria were definitely established for each of the modes of apperception, the responses on the lists were scored. First the class-divisions were scored in red. Thus in Table 3, scores for *"one person," "two persons," "two persons"* card reversed, *"one animal," "animal skin," "two animals"* and *"flying animals"* appear in red. Then each response was scored *W, D, Dd, Do,* or *DS.*

B. *Scoring the quality of apperception.*

1. *Form (F).*

Rorschach distinguished between forms sharply perceived ($F+$) and forms poorly perceived ($F-$) on a statistical basis. According to Rorschach, a good form perception is a response given by a majority of healthy adults. $F+$ is held to be a criterion of clearness or keenness of perception. Rorschach makes it depend upon fre-

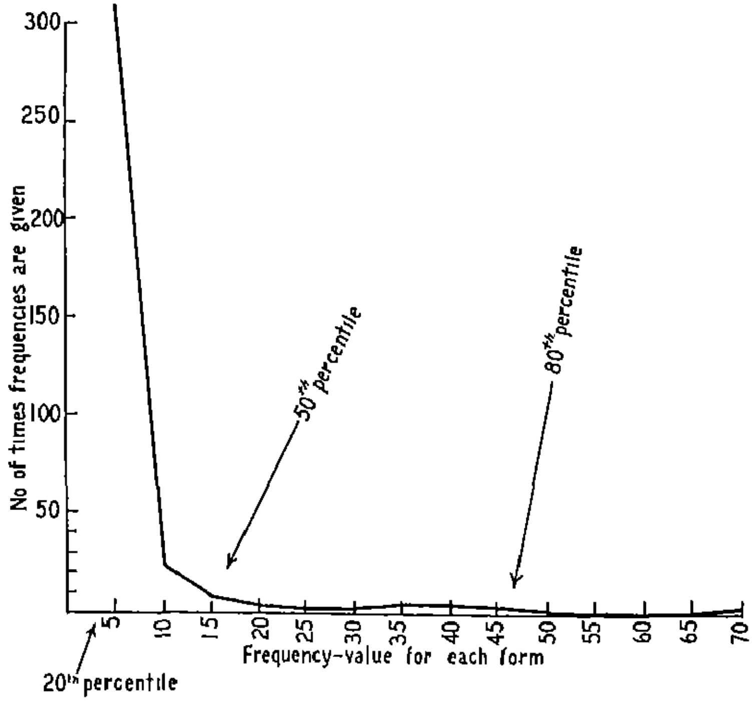


FIGURE 3

CARD II — FREQUENCY WITH WHICH THE SAME OR SIMILAR FORMS ARE GIVEN

quency, implying that the frequency of the form shows whether that form corresponds to a clear perception of a majority of normal adults. According to Rorschach, the forms given most frequently by normal subjects are good forms and are considered "standards" for judging other forms Those that are better than these "standards" are designated plus and those which are poorer, minus Thus this designation is given partly as a result of empirical and statisical procedure, partly because of the judgment of the examiner (25)

Rorschach frequently admits that subjective estimate must enter in the determination of the good and poor forms. In fact observation of the scoring of this factor in his manual discloses many inconsistencies Löpfe (17) argues that only after long practise with the test and only after comparison with the data of other investigators and especially with Rorschach's records, will it be possible to attain sufficient objectivity in judgment. According to Loosli-Usteri (16) on the other hand, subjective estimate plays only a small role. Oeser (23) seems reconciled to the fact that a certain degree of subjectivity is inevitable in scoring *F*—. No hard and fast rules can be laid down.

> There are a certain number of forms on each plate that are seen by nearly everyone. These are called *F*+. Others have to be judged on their individual merits. On the whole, experience shows whether the form is really vague or far-fetched. One can always question the subjects and so come to a decision (23).

Beck (3) appreciates the difficulty in scoring *F*+ and *F*— without a standard list and reports glaring inconsistencies observed in the work in Zurich Meltzer (19) reports that he scores *F*+ in the light of a frequency distribution of form responses. His lists have not been published, however. Gardner (8) arbitrarily determined that a blot or portion given the same form five times (in responses of 100 subjects) was *F*+; all others were scored *F* without designation

It is obvious that there is still much subjective estimate involved in scoring the form answers and there is urgent need for a standard list of good forms. Such a list was prepared in the Brush investigation for the age group studied, based on the frequency distribution.

As indicated above, Column XI of the preliminary lists gives the "frequency criterion" since it summarized the total number of times references had been made to certain forms. It will be recalled that the responses had been arranged according to the nature of the content, and further, according to a preconceived classification dependent upon similarity of form On the lists, the frequencies for the form-classes had been summarized in red. This red figure, therefore, in Column XI indicated the frequency with which particular forms had recurred and thus was taken as the "frequency value" for the forms (see Table 3).

Frequency values were arranged in a distribution and graphs were made and studied for each blot in order to locate that point where a break seemed to take place, separating the most frequently perceived forms and those less frequently perceived. Figures 3 and 4 present graphs for two of the ink-blots. Percentiles were located for each distribution and frequencies and percentiles summarized

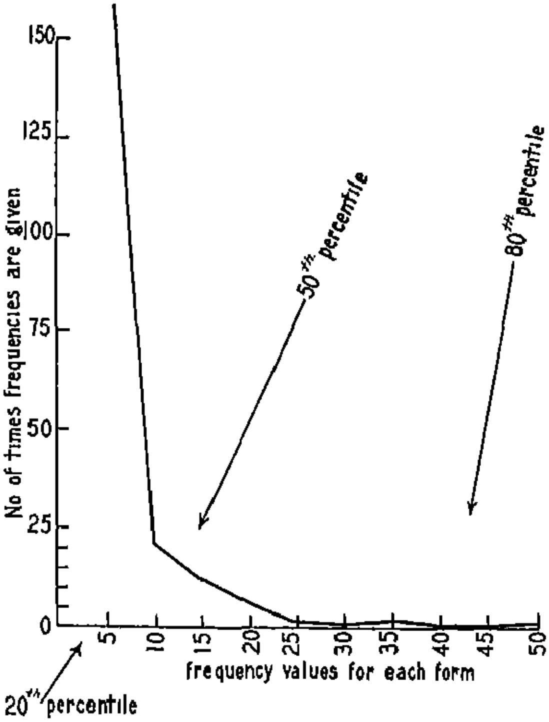


FIGURE 4
CARD VIII — FREQUENCY WITH WHICH THE SAME OR SIMILAR FORMS ARE GIVEN

on a separate sheet. All frequencies that fell at the 50th percentile were considered good forms. For most blots, the median seemed to fall between 11 and 15. Therefore all frequencies of 13 or more were considered good form ($F+$).

Forms which did not receive a high frequency had to be judged as "better" or "worse" than the "norms." Often responses were given which had a low numerical frequency but which were obviously like the good forms. For example, the responses were expressed in different words, as *"scene," "landscape," "mountains and forests in the distance,"* or *"panorama."* In order to score these forms in an objective manner, another criterion was adopted.

Groupings were made for all forms of like kind given as a whole

TABLE VI

APPENDIX: CARD V—GROUPINGS OF THE SAME OR SIMILAR FORM IMAGERY

Note	Symbol	Response	Column IX	X	XI	XIII
1		One person seen in center				F+*
	W	person	11		11	F+
	C	person	2	(11)	13	F+
	1	head, face	2	(2)	4	F+
	1 (F) C	head, face	8		8	F+
	5	ears	1	(1)	2	F+
	2	legs	2	(4)	6	F+
	S	costume		(7)	7	F+
2		Two persons seen				F—*
	W	persons	6		6	F—
	S	persons	1	(6)	7	F—
	6 & 7	feet	6	(2)	8	F—
3		One animal seen, stretched out				F+*
	W	animal stretched out	5		5	F+
	C	animal	3	(2)	5	F+
	1	head of an animal	4		4	F+
	5	ears of an animal	3		3	F+
	5	horns	2	(3)	5	F+
	5	legs		(1)	1	F+
	2	ears	3		3	F+
	2	feet	8	(9)	17	F+
4		Two animals				F—*
	W	two animals seen together	6		6	F—
	S	animal	2	(4)	6	F—
	S	part of an animal	4		4	F+
	S	animals, without heads	1	(1)	2	F—
	6	feet	16		16	F+
	7	feet	6	(1)	7	F+
	6 & 7	feet of animals	6	(2)	8	F+
	S	dogs	3	(2)	5	F—
	S	animal's body	2		2	F+
5		flying animal				F+*
	W	flying animal	3		3	F+
	W	bat, bird, butterfly	118		118	F+
	W	wings	2		2	F+
	W	insect of some kind	255		255	F+
	C	body	2	(2)	4	F+
	1	butterfly's head		(1)	1	F+
	5	feelers	1	(5)	6	F+
	2	feelers	3	(2)	5	F+
	S	wings	4	(18)	22	F+
6		Two flying animals, card reversed				F—*
	W	two birds	2	(4)	6	F—
	2	bird's heads	4		4	F—
	S	Part of a chicken	1		1	F—

*Score written in red, showing it is applicable to the entire group unless otherwise indicated *or in any of their details* with a summary of the frequency of each. These were added to the preliminary lists in the form of an appendix, reference being made to them in the lists in blue ink in Column IV. Table 6 presents a page from the appendix to the list for Card V. By consulting these groupings, it might be determined whether a form was + or — despite the fact that its numerical frequency was lower than that which the numerical criterion demanded. For example, *"flying animal"* given only three times would be *F*— according to the numerical criterion. Yet *"bat," "bird"* and *"butterfly"* were given 118 times and insect, 225 times, showing that *"flying animal"* is plus.

It was likewise observed that despite the use of the two criteria already referred to, there were some few responses which could not be scored objectively. They were responses which had a low numerical frequency and which could not be compared with any of the groupings in the appendices. Subjective estimate had to determine the quality of these responses. For example, Card V, *"an ornamental dish on a stand,"* or Card VIII, the small space forms in the center of the card, *"a milk bottle"* were responses given only twice each, yet were considered *F*+.

In order to reduce the error involved in such subjective procedure as much as possible, the scoring was done by three judges, independently. The examiner scored each response in the lists in the appropriate column (Column XIII). The left side of the lists were turned down once, and another judge scored them. The sheet was again folded over and a third judge scored the responses When the judgments were compared and discrepancies occurred, two other judges were consulted before the final score for form was decided upon.

Summary for the determination of F+ and F—.

Three criteria were used for the determination of good form:

1. All forms which received a frequency of 13 or more were considered good forms and scored *F*+.
2. All forms which received a low frequency but which belonged to a similar group or class were scored as that group was scored.
3. All forms of low numerical frequency, where there was no basis for comparison with the same or similar forms, were scored according to the concurrent estimates of from three to five judges.

2 *Movement Responses* (*M*).

The movement factor is one of the most important in the test; it is also the most difficult to score. Rorschach (25) advises that movement answers are "engrams of movement in prior experience, seen, imagined or carried out by the subject himself." The blot or parts of the blot must be seen as if in movement. The subject must feel the object to be in motion. In Card I, for example, *"two forms of angels with fluttering wings,"* and Card II, *"two fools dancing together"* are movement responses. Often, the subject himself explains that he experiences movement. He may even show by his actions or conduct that movement has been seen and felt.

Rorschach specifies that "movement occurs almost always with human beings or animals having human characteristics." However, on occasions, there may be movement with animals or even lifeless forms,—plants, objects, even geometric figures and isolated lines. Further, Rorschach's movement factor is rarely found with small details. If movement is expressed, it is merely a secondary embellishment.

Rorschach cautions that care must be taken not to view every movement that is mentioned as a movement response. Any response in which an active verb appears in the present tense need not be a movement response. Answers such as *"a duck that goes into the water," "a dog which snaps at a butterfly," "a flying bird," "an aeroplane in flight,"* or *"a mountain spitting forth fire"* in most cases are not movement answers according to Rorschach. They are form answers, the interpretations being based on the form of the parts alone, and the statement of movement is often only some secondary association.

To score an answer as movement, the examiner must be assured, then, that a primary feeling of movement really determined that particular response. The subject must feel motion in some way. The movement must appear to be actually executed, so that there appears to be a reciprocal change of position of the parts of the blot. As Beck (3) explains, "there is represented in the response an actual experience, in which the individual himself is, in the fantasy, engaging in some activity, conscious or unconscious." Thus, *"two men fight"* is a movement response only if the subject actually experiences the movement. *"A bandmaster"* for Card III, reversed, is a movement answer, because the form of the blot alone could not have produced that response. Movement was seen, the hands stretched up in the air, as if leading a band.

Even if no actual movement takes place, but the response implies strain or potential movement, it is considered a movement response. *"A man lying down or standing up"* is, for example, a movement response.

Some investigators do not adhere to this orthodox view of the movement factor. They score as movement, any response in which movement is indicated, whether it be a human being, an animal or object. As a result, much confusion exists in scoring this factor and many discrepancies may be observed in the literature on the subject Card VIII, for example, is frequently interpreted as "an animal climbing over rocks or going up a tree." In some records it is scored *form,* in others, *movement.* Perusal of Loosli-Usteri's records (16) shows that she restricts herself to Rorschach's original specifications, scoring many responses which express movement verbally as form answers. Kerr (14) on the other hand, does not adhere to the orthodox view and scores many responses where movement is expressed, as movement.

As Beck (3) so ably points out, the scoring of this factor is very important since it is one of the most fundamental elements in the whole Rorschach system of personality diagnosis. There must be some unanimity of opinion in scoring it, if the whole system is to mean anything.

Some investigators assume that intelligent subjects can tell whether movement has been involved or not. They question their subjects cautiously. Unfortunately, the subject cannot always help because he is not always aware of the experience of movement Loosli-Usteri (16) reports that she scored many responses as movement, even when the subject denied that movement had influenced his answer. She did not feel that her young subjects were capable of introspection Beck (3) also points out that the movement may be unconscious Questioning may elicit answers which are merely after-thoughts, not really what actually was felt at the time the answer was formulated. For this reason, he does not trust the subject's introspection He again suggests research on this subject, to determine statistically, what movement responses are generally given. Levy (3) on the other hand, does not find this difficulty in determining "felt movement" in Rorschach's sense. Seeing of animals in movement, he claims, does not represent "felt" movement; a bird with wings outspread is not "felt" movement, but men moving or lifting up something is

In some cases, doubtful answers may be compared with others to see whether *F* or *M* predominates in the whole record. "In cases in which it is questionable whether a mere form response or a combined response is present, nothing further can be done than to write *F* for the time being, and to correct it, later perhaps, after a comparison with all the responses given" (26). It is evident that with this procedure, subjective estimate may enter to such a degree that it may determine the whole psychogram. Beck's objection (3) to this manner of scoring certainly seems to be justified. Levy (3) on the other hand approves of this common practise. If a movement tendency is observed in a subject's results, he is willing to conclude that the doubtful responses have been influenced by movement In fact, he prefers this procedure to Beck's suggestion for some statistical determination of movement responses (3).

More research is obviously necessary to determine upon adequate criteria for scoring the movement responses. In the absence of "norms" or "standards," the following criteria were used in determining the movement factor in the records of the Brush investigation:

1. Responses were scored movement where the language indicated "felt" movement in a human figure.
2. When movement was expressed with other responses—with animals, natural forms, or objects—comments of the examiner which were written on the record, answers to cautious questions, and explanatory remarks of the subjects determined whether or not movement had been involved.
3. Where change of support or position of parts of the blot was clearly indicated, movement was inferred.
4. Where the subject indicated by word or gesture or both that he actually experienced movement, strain, or tension, movement was indicated.
5. The general criterion applied was,—would this answer have been possible if movement influences had not been operating.
6. In a record where there were numerous movement responses which could not be challenged, a movement tendency was assumed and doubtful responses were scored movement.

7. From experience, some responses were recognized as movement answers generally given by subjects and scored as such by Rorschach and by subsequent investigators. If the subject did not actually express the movement, cautious questions were asked in order to verify the influence of movement.

Experience with "normal" groups or subjects has impressed the writer with the value of having the subject help as much as possible in determining the factors involved in a response. By word or gesture, he frequently indicates whether movement or only form was instrumental in the interpretation given. Occasionally, this is not possible No doubt with subnormal or certain mentally disordered subjects such procedure is never possible [9]

3. *Color Responses* (*C*)

It was not difficult to determine whether a color answer had been given when color was expressed. *"A pink pig," "green water,"* and *"colored butterfly,"* were answers which showed at once that color had influenced the interpretation to some degree.

---

[9] At present, the Brush Foundation is experimenting with finer differentiations of this movement factor Movement responses are scored in three ways:

M refers to movement of human or animal forms according to the criteria specified above

(*M*) refers to any implied movement, strain or tension

m refers to movement in nature, natural or artificial happenings It seemed necessary to elaborate upon Rorschach's original movement factor to include other forms of movement because such factors seem to have diagnostic significance.

In a recent conversation with Dr. Bruno Klopfer (New York City), he reported that he was making further differentiations in this movement category, distinguishing between actual movement, movement tendency, "dynamics in happenings," and certain kinds of shaded responses closely allied with the movement factor

In the last issue of the *Rorschach Research Exchange*, Klopfer and Sender (34) differentiate between *M* or Movement of human beings, *FM* or Form tending toward movement which appear to include human and animal forms, and *m* or minor movement, this last referring to *a*) expressive qualities in parts of human and animal forms as "grinning faces" or "pointing fingers" and *b*) passive happenings as "explosions, collisions, falling objects."

The writer has also had correspondence with Mary Hunter from Honolulu, in which she states that she is of the opinion that it will be necessary to make additional differentiations in this movement factor, possibly *M*, *FM* and *MF* to indicate the degree of *F* included with *M* as in the case of the color answers.

Finally, in a recent note by Thornton, G R (36), he presents statistical evidence which points to the conclusion that a better measure of movement is obtained by including all responses which otherwise correspond to Rorschach's concept of movement response, whether their content is human or non-human. This is in accordance with the suggestions contained in the criteria cited above.

If, however, no color was indicated, the examiner had to determine whether or not there were any color influences. Judgment had to be used, also, in making the three color differentiations. Was it pure color that influenced the answer? Or was it both form and color, the form being predominant *(FC)*? Or was it more color than form *(CF)*?

Pure color is, of course, easy to recognize. *"This is water, because it is green," "this is Russia because it is red on my map,"* and *"jello"* are obviously pure color answers.

FC and *CF* are not as readily distinguished. As already indicated in a previous paper (Note 2), proper administration of the test is extremely helpful in the scoring of these responses. The examiner must get as much assistance from the subject as possible at the time of the examination. Rorschach permitted questioning the subject after the test was completed Many examiners follow him in this manner In the Brush investigation, however, it was found more effective to ask "safe" questions after each card was returned rather than to wait until the end of the test for all questions Frequently a subject gave from 40 to 50 responses. He was apt to forget by the end of the examination what had really influenced a particular response, color or form.

In this experiment, every care was taken to make these decisions during the examination. Cautious questioning almost always gave the necessary information Of course care was taken not to suggest the classifications in these inquiries If these "safe" questions did not elicit adequate information, further questions were asked at the end of the examination. These had to be more direct. Rorschach himself suggested one question *"Would this look like a butterfly to you if it were not red?"* Beck (1) thinks that all questions should be indirect. Vernon (30) on the other hand, sees no harm in asking the subject directly whether color was involved or not. In this investigation, only indirect questions were resorted to. Questions of this nature were asked: *"What made you say this looked like a butterfly?" "If it were like this* (pointing to the black), *would it still be a butterfly?" "Suppose it were like this,* (pointing to another color) *would it still be a butterfly?"*

As with movement, some investigators depend upon the tendencies displayed by the subject in the rest of the record Loosli-Usteri (16) applied the general rule of assuming that if a record contained many primary color responses *(C)*, it was more probable that a doubtful response was a *CF* than a *FC* Rorschach permitted this reference to the record for the doubtful answers (25). In this experiment, only when the rest of the record was heavily weighted with color answers, *C* and *CF*, was it assumed that the doubtful answers were *CF* rather than *FC*.

Finally, many investigators rely upon their experience and their observation of how other examiners score similar responses in literature. Loosly-Usteri (16) for example, admits that in most cases, she proceeds by analogy. In this connection, Beck (3) suggests another research problem, i.e., determining statistically, tables for these various color categories and deciding experimentally which procedure is best, introspection or frequency tables.

In this experiment, then, the examiner depended (1) upon explanations given by the subject either voluntarily or in answer to questions, (2) on the color tendencies displayed by the subject in the rest of the record, and (3) upon her own experience and observation. Frequently the subject explained gratuitously *"The color is there although I suppose the form isn't so good," "This is the sunset because the rays shoot out and it is the color of the sun"* for Card II, *"these flags are hung to a pole, they seem all torn and weather beaten"* for Card VIII, *"these are blue pillows, not much the shape of pillows, I guess the blue reminded me of our pillows at home,"* Card VIII. Such explanations helped immeasurably in determining the color score

Of course, sometimes the subject could not account for his answer At other times, he would insist upon color only when the experience of the examiner made her relatively certain that some form was involved. For example, an answer of *"sunset"* for the red bottom, Card II, inverted, was explained by one subject as *"the sun because it is all red, no, it is not the form of the sun"* Yet this same subject did not interpret the other red portion on the same card, as *"the sun"* Again, a subject would say that color did not enter at all in determining his answer, when it appeared obvious that it did. For example, Card IX *"this is sand* (top brown), *this is water* (middle green) *and this is the sun, I suppose,* (bottom red)" was explained in this fashion, *"sand is sometimes piled high in a mound and can come to a point, and you can have round puddles of water."* For the *"sun,"* all he would say, *"You could have the sun that way."* It was obvious here, that color was much more influential and that the subject was just trying to justify his answer by fitting in some form. The examiner must, therefore, rely to some extent upon his experience with the test and must use his judgment and discrimination.

Color scores may be good or bad, according to whether the forms are good or bad. Rorschach did not score color answers + or —. He assumed that *FC* meant good forms and *CF*, poor forms. Experience with responses, however, shows that frequently even when color predominates, the form is good so that one may have *CF*+ and *CF*— Generally *FC* is a good form. But, even here, form may predominate and color enter slightly, and yet the form be poor. For example, a reply of *"snout of a dog"* for the red *"butterfly"* figure, center, Card III, because *"it looks like a dog is coming toward you and this is shaped like his nose,—and it is reddish, too, you must admit"* was scored *FC*— because the form seemed to dominate the answer and color entered admittedly, but the form (according to the Frequency Tables) is a poor one. Again for Card VIII, the red-orange figure at the bottom, the answer *"butterfly"* which was most common, was scored *FC*+ or *CF*+ depending upon the emphasis given by the subject.

As with movement, the color scores were not introduced in the Frequency Tables because, as indicated above, the examiner had to make the determination in reference to each individual. However, reference had to be made to these tables to determine the + or — score in the color combinations.

4 *Chiaroscure answers* [*F*(*C*)].

Responses to the black-gray parts of the blots were designated *F(C)*. These were readily recognized. The subject generally indicated when giving the answer or in reply to cautious questioning, that the shaded parts had stimulated the response. *"This is chinchilla because it is shaded that way;" "This is a scene as if you are looking at it from a distance, the darker part being trees, the lighter clouds," "these are hazy mountains as seen from the distance"; "This is a face, you can see in the dark and lighter spots, the features of a face,"*—were scored *F*(*C*). For *"clouds"* or *"smoke"* or *"gran-*

it," cautious questions generally brought out the fact that the shadings had influenced these responses and hence the $F(C)$ score was applied. No finer differentiations were made for the chiaroscure responses in this investigation.[10]

5. *Combinations*

At times, movement, colors, and shadings may all be involved together in different combination in determining the response given Rorschach included such cases where movement and color both influenced the answer, but claimed that this occurred infrequently. Upon closer examination, he observed, either movement or color was primary, the other secondary However, where both seemed to occur simultaneously as with highly gifted individuals, for example, he designated the score *"MC"* or *"CM"* according to the answer (25)

In the Brush investigation, when it was clear that such combinations existed, the various categories were indicated. For example, Card X, *"View of the bottom of the sea at Bermuda You can see the colored vegetation and different kinds of tropical fish swimming around, some pulling and clawing each other"* was scored *FC*+ and *M* because these factors seemed to be equally influential in determining that answer.

C *Scoring the content of the apperception*

Responses were described according to the nature of the content in Column XIV of the Frequency Tables with abbreviations used according to a prearranged plan. Reference should be made to Table 1 for a list of these abbreviations

D *Originality (O) and popularity (P) of response.*

1. *Originality (O)*

Rorschach's criterion for originality was that an answer occurred not oftener than once in 100 times. He differentiated between good original (O+) and poor original answers (O—) according to whether the form, movement, or form-color was good or poor. He computed the percentage of original answers and indicated by "±" or "∓" which was the most frequent, the + or — (25).

---

[10] Vernon (30) suggested that the shaded response be further differentiated into *F Ch*, *Ch F*, and *Ch* according to the emphasis placed on the form and the shadings and that a sum *Ch* be obtained just as with the color score. Guirdham (9) followed his suggestion in his study with epileptic subjects This category was studied in detail by Binder (5) who suggests further distinctions: *"F (Fb), Hd, F Hd and Hd F"* because they appeared to have special diagnostic values Following him, Ganz and Loosli-Usteri (7) score for these different types of shaded response Klopfer claims to have arrived independently at similar subdivisions of this category. The Brush Foundation is now tentatively using the following. *F (Ch)*, *(Ch) F*, *and (Ch)*.

Experience with this category shows that it is not so simple to score as Rorschach believed Complexity of response and variability in language cause difficulty in determining originality. When is an answer one in 100—when the forms are unique? when the words are unusual? when different settings are given to forms usually given? Many questions come up when use is made of this numerical criterion. When, for example, an answer contains a form which is usually given but a different name is applied, the answer cannot be considered original An answer of *"cobra"* given once to a detail for which the response *"snake"* appears hundreds of times surely is not original

Most investigators determine originality subjectively. Beck (1) reported in an earlier paper that in scoring for originality, he followed the precedents established by Rorschach, Oberholzer and Levy. But on innumerable occasions, he had to rely on his own subjective estimate. In a later paper (3) however, he admits he has discarded this category because it offered nothing he could rely upon with any degree of certainty. Gardner (8), too, disregarded this factor, as have many other Rorschach examiners.

Few standards of originality are available in published form. It is obvious that even if they were, they would not be applicable to every group. What is original for one age-group in one locality may not be for another

Vernon (30) suggested an index combining both the original and the popular factors, an "originality-commonality index" as he called it. Tables should be prepared of responses from a large sample, several investigators determining a rating for each response in reference to goodness or poorness of originality Meltzer (19) determined upon a method which he claims made it comparatively easy to differentiate O+ from O— He introduced a new measure (O-P) as a substitute for the O and P which he considered most inadequate. This index is a measure obtained by assigning values based on a study of the frequency distributions of all responses, normal details as well as wholes Frequencies of 15 or over are assigned the value of

1, 6 to 14, 2, 3 to 5, 3, 2 the value 4, and 3 to 5, the value, 3, 2 the value, 4, and 1 the value, 5.

In the Brush investigation, the original answers were determined statistically as the form answers had been, using the same frequency distributions and the same graphs. Just as the frequency criterion for good form developed around the 50th percentile, the 20th percentile was used for the line of division between the usual forms and the unusual answers (see Figures 3 and 4).

All forms with frequencies that fell below the 20th percentile were considered original. With eight test cards, this percentile fell below four, with one, below five, and with one, below six. It was decided to consider all forms given three times or under (in 300 subjects) original forms. This corresponded with Rorschach's criterion of "one in one hundred."

Responses were scored *original* when

1. An ordinary form was given but in a setting determined to be unique by the frequency, one in 100. For example, Card I was interpreted as *"Two men on each side clinging to each other. This must be an acrobat act, in a show or probably at the circus."* Frequently the side figures were interpreted as persons" and the middle figure as "a person," but the combination was sufficiently unique to class the whole response as an original.
2. An ordinary setting was given but the forms involved determined to be unique by the frequency, one in 100. For example for Card X, an answer sometimes obtained was *"some scene under water."* The response *"A scene in Bermuda, as seen through the glass bottom of a boat, showing the colored coral reefs and the colored vegetation, and the various tropical fish"* contained forms sufficiently different to warrant the score, original.
3. An entirely unique configuration was given. For example, for Card I, *"northwind puffing and blowing"* was considered original as was *"Test-tube"* for Card IX, center space form, and *"Gaseous substances"* for the entire card.

Original answers were scored good or poor (+ or —) according to the quality of the form or forms included.

Highly original answers which appeared only once in the entire group were scored *Individual* (*I*). An individual answer was the only one of its kind which appeared.

At times, answers were scored original according to the above criteria, which obviously depended upon the subject's home environment, training or specialization. Numerous anatomical forms, for example, from a child whose father was a professor of anatomy, had to be scored original These facts had to be taken into consideration, however, in the interpretation of the scores Obviously, such responses were not indicative of either *originality* or *eccentricity*.

2 *Popularity or Commonness of Response (P).*

Rorschach (26) suggested that a popular response was one given by "one in three persons" Other examiners, as Löpfe (17), Loosli-Usteri (16), Kerr (14), Guirdham (9), and Gans (7), use the criterion one in six Most studies seem to take this latter criterion for their popular forms. Linares Maza (15) on the other hand, uses one in five, Vernon (30), one in three. Gardner (8) arbitrarily scored *P* all responses given 25 or more times by his 100 subjects. Meltzer (19) discards the category as already indicated and uses a new measure, the (*O* — *P* index).

It has frequently been observed that a response may contain a popular form, yet the entire response may not be popular. For example, in Card IV, a "*boot*" or a "*foot*" is popular, yet the response, "*a deformed man with tiny head, crooked arms and tremendous feet*" is not a popular response, strictly speaking. Again, for Card III, "*Two men*" is a popular response, but "*Two men mixing potents in a cauldron*" is not only not popular, but is original.

In attempting to establish a criterion for this test factor, the Brush investigators resorted to the distribution for the form frequency. Studying the graphs, the 80th percentile appeared to be a good criterion for the popular response (see Figures 3 and 4).

With the test cards, the 80th percentile fell between 33 and 118 The percentile values were averaged and the result 48.2 was taken for the criterion for popular answers. Therefore all forms given approximately 50 times were scored popular. Summaries were made of the popular responses for each card and for the entire set. Table 7 summarizes the popular responses for each test card as determined by the criterion adopted.

Using the criteria determined upon for the originality and popu-

TABLE 7

SUMMARY OF POPULAR FORMS AS DETERMINED BY 80TH PERCENTILE CRITERION

Card		Location*	"Popular" Form
I	W	Whole figure	face
	W	Whole figure	bat, butterfly, moth, bird
	S	Side figure	person
II	W	Whole figure	butterfly
	S or 5	Side figure or black figure at sides	bear, rabbit, dog
	S	Side figure	man, person
III	W	Whole figure	persons doing something as picking, carrying, pulling
	5	Middle red figure	butterfly or moth
	2	Side "leg" figure	leg or foot
IV	W	Whole figure	animal skin, hide, fur, pelt
	W	Whole figure	bat, butterfly, fly, moth
	1	"Foot" figure	foot, boot or shoe
V	W	Whole figure	bat, butterfly, moth, bird
VI	W	Whole figure	animal skin, fur, pelt, rug
	S+S	Lower two-thirds	animal skin, fur, pelt, rug
VII	1	Top third of side	face or head of person
	2	Middle third of side	face or head of animal or person
	3+3	Bottom third of whole	butterfly, moth
VIII	1	"Animal" figures at sides	four legged animal of some kind
	4+4	Top middle gray	tree, tree parts
	3+3 or 20	"Rib-figure" with or without rest of blue	spine, ribs, fishbone
	2+5	Red-orange bottom	butterfly
IX	3 or 13	Bottom, pink, entire pink or just outer portion	person's face, head
	10	"Animal's face" in green on border of brown	animal's head
	2	Green middle	person bending over
	D3+C+3	Pink with center, card reversed	tree
X	12	"Rabbit's face" figure	a rabbit's face
	V20	Card reversed, lighter green center of green middle	a man, a person
	6	Blue figures, sides	crab, spider
	8	Orange figures, sides	dog, lion
	3	Green figures, center	worms, caterpillars, sea horses

*See note at bottom of Table 5.

larity of response, the answers were scored by the three judges, final determination made for each, and entered in Column XV. As with the forms, where disagreement occurred among the three judges, two others were consulted.

It should be mentioned that not infrequently a popular form was included in an answer which in itself was not popular; it might even occur in an original answer. In these cases, the score was enclosed in parentheses thus, (*P*), meaning a popular form is included in the response.

Responses not classified as popular or original belong, of course, to the intermediate range and have no designation in the column.

Summary of the evaluation of the test scores

1. Preliminary lists were constructed in which every response, every item in a response, and all classes of responses (grouped for similarity of form) appeared with their respective frequencies.

2. Definite criteria were determined upon for certain of the test factors, based on the frequency values found in the Preliminary Lists.
   - *a* The 20th percentile was selected as the criteria to distinguish *normal* from *rare* details. All details with a frequency above the 20th percentile were considered *normal* details.
   - *b.* Keenness of *form* perception was based on three criteria,—(1) a numerical frequency of thirteen or more; (2) similarity with groups of usual forms; (3) subjective estimate of three to five judges in cases where a numerical criterion was not applicable.
   - *c* *Originality* of response was based on three criteria,—(1) a usual form in an unique setting; (2) a unique setting with usual forms, and (3) an unusual configuration without any usual forms, in all three cases uniqueness being determined by the frequency one in one hundred.
   - *d.* All forms given 50 times or more were considered *common* or *popular* forms.

3. Qualitative criteria were suggested for use in scoring other test categories.
   - *a.* *Whole answers* were determined from the responses as recorded on the original record blanks.
   - *b* *Space details* were those responses which could not have been given, if the spaces had not been selected for interpretation.

c. *Oligophrenic details* were determined by consulting the lists for the frequency with which the *whole* form and *part* forms were given.

d. *Color scores* had to be determined from the responses and notations as recorded on the original record blanks

e. Certain criteria were suggested for aid in determining the *movement* answers These were applied to the answers as they appeared on the original record blanks.

(1) An answer where the language suggested movement in a human figure was considered movement when the answer could not have been given if movement influences had not been operating.

(2) Movement expressed with other forms, not human, was studied in the light of the subject's remarks and answers to questions and observations and comments of the examiner

(3) There had to be evidence that the subject had experienced the movement.

(4) Comparison with other answers of the subject disclosed whether there was a tendency toward kinesthesis

IV THE COMPLETION OF THE FREQUENCY TABLES

After each response on the preliminary lists had been scored by the three judges, or in cases of disagreement, by the five judges, the resulting scores were copied on the index cards in the catalogues The next task was to arrange the responses in Frequency Tables in such manner that they might be applied practically in scoring the test records

After much preliminary experimentation, which included arranging responses alphabetically without considering location of detail, in groups corresponding to the details of the blots, and in groups according to the nature of the content of the response, it was decided to combine all these procedures Responses were divided into groups corresponding to the *gross divisions* of the blots Within each of these divisions, responses were grouped according to the *nature of their content,* into the following general classification:

I Animate forms
  A. Human forms
  B. Animal forms (not human)

II. Anatomical forms
III Inanimate forms
  A. Natural forms
  B. Artificial forms

Finally, within these subdivisions, responses were arranged alphabetically. Frequency Tables were prepared in this manner for the ten blots. From three to twelve lists were allotted to each ink-blot Forty-six lists were prepared in all.

TABLE 8

SAMPLE PAGES TAKEN FROM FREQUENCY TABLES

CARD I. FREQUENCY TABLES DIVIDED INTO 3 LISTS

LIST A. Responses to the *WHOLE FIGURE* and to the *ENTIRE SIDE FIGURE*

W	whole figure
S	side figure
1/3 up	upper third of figure
1/2 up	upper half of entire figure

LIST B: Responses to the *DETAILS IN THE SIDE FIGURE*

21	upper third of side figure
10	lower two-thirds of side figure
2	"wing" projections of side figure
12	"head" part of side figure
11	profile along outer edge of "head" part
29	small projection along bottom edge of side figure
18	profile, under edge of side figure
6	space form, upper level
7	space forms, lower level
4	projections of side figure, joining it with middle
5	projections bottom of side figure, joining it with middle

LIST C Responses to the *MIDDLE FIGURE* and to *DETAILS IN THE MIDDLE FIGURE*

C	middle or center figure
17	middle figure, shaded part only
31	upper third of middle figure
1	claw-projections, top middle, outer projections
9	claw-projections, top middle, inner projections
15	lower 2/3 of middle figure
3	bottom center projection
16	space form between middle claws
33	space form between middle and outer claws
8	space form, center of figure
40	large space form, above entire middle of figure

TABLE 8 (*continued*)
SAMPLE PAGES TAKEN FROM FREQUENCY TABLES

(human)
I A

CARD I LIST A: Responses to the *WHOLE FIGURE* and to the *ENTIRE SIDE FIGURES*

1 Pos	2 Response	3 Loc	4*	5 W,D	6 F	7 Content	8 O,P
	I. *ANIMATE FORMS*						
	A *HUMAN FORMS*						
AV	acrobats	W or S		W or D	F+	H	O+
	angel (body in center & wings on sides)	W		W	F+	(H)	
	angel without a head	W		W	F+	(H)	
	angels, (two)	S		D	F+	(H)	
	angels with wings on them	S		D	F+	(H)	
	angels with wings, together	W		W	F+	(H)	
	children (see "PERSONS")	S		D	F+	H	(P)
	clown's head	W		W(DS)	F+	Hd	
	dancers	S		D	F+	H	(P)
	dancers kicking or hanging from ceiling	W		W	F+	H	O+
	devil	W		W	F+	(H)	
	devil's head	W		W(DS)	F+	(Hd)	
A>	fairy because of wings	S		D	F+	(H)	
	fairy because of wings at sides	W		W	F+	(H)	
	girls dancing (see "PERSON")	S		D	F+	H	(P)
	girls dancing, holding hands, going around in circle	W		W	F+	H	(P)
	goblin	W		W	F+	(H)	
V	god's face, face of some god	W		W(DS)	F+	(H)	
	human beings (see "PERSON")	S		D	F+	H	P
	human pyramid made up of acrobats	W		W	F+	H	O+
	hunter	S		D	F+	H	
	hunters taking animal away	W		W	F+	H	O+

TABLE 8 (*continued*)

SAMPLE PAGES TAKEN FROM FREQUENCY TABLES

(human)

I A

CARD I LIST A. Responses to the *WHOLE FIGURE* and to the *ENTIRE SIDE FIGURES*

1	2	3	4*	5	6	7	8
Pos.	Response	Loc		W,D	F	Content	O,P
	lady (see "PERSON")	W		W	F+	H	
	lady with wings	W		W	F+	H	
	ladies (see "PERSONS")	S		D	F+	H	P
	ladies with wings, with capes flowing, with robes	S		D	F+	H	(P)
	ladies with wings, with capes, together, dancing	W		W	F+	H	(P)
	ladies in the center holding men's hands	W		W	F+	H	
	man (see "PERSON")	W		W	F+	H	

*Column 4 has been omitted

For example, Card I was divided into three divisions, the whole figure, the side figures and the center between these sides. Therefore the responses were divided into three lists, the first including all answers to the whole figure and to the entire side figure, the second, answers to details in the side figure, and the third, responses to the center part. On the other hand, responses to Card X were divided into twelve lists, corresponding for the most part to the different colored portions.

It should be mentioned that responses were frequently classified not only in their proper alphabetical location, but also under class designations, such as botanical, structural, water, fire, geographical forms, and so forth. In this way, every effort was made to facilitate locating a response with the greatest possible speed. "*Christmas tree*" for example, was tabulated under "*CHRISTMAS TREE*," under "*TREE*," and also under "*Botanical forms*." Accordingly, many cross-tabulations appear. This was done not only to assist in locating responses rapidly, but to make the tables suitable for other test material. Inasmuch as the tables do not exhaust all normal possibilities of response, class designations and cross-tabulations will assist in the scoring of new verbal responses which do not appear in the lists.

Frequency Tables with their directions for use were thus assembled. Table 8 presents sample pages taken from these lists showing top sheet for Card I, and the first page of responses. It will be noted that each table is arranged in eight columns, as follow.

Column I	Symbol representing position of card
Column II	Response
Column III	Symbols showing location on blot
Column IV	Description of location on blot
Column V	Score which has been determined upon for *W*, *dW*, *D*, *Dr*, *DS*, and *Do*
Column VI	Score for good (+) and poor (—) form
Column VII	Indication of the nature of the content
Column VIII	Score for originality (*O*) and popularity (*P*)

Each page is thus appropriately labelled and the tables for each test card are assembled with a preliminary survey giving the necessary explanations.[11]

The Frequency Tables in their present form represent the standard of normality for the specific age-group and for the particular sample studied. Their use is to determine the score for the responses in respect to the following test factors: *W*, *D*, *Dr*, *Do*, *F*+, *F*—, *O*+, *O*—, *I*, *P*.

V. SCORING TEST RECORDS USING THE FREQUENCY TABLES

The records of the Patrick Henry Junior High School subjects could now be scored. By means of the tables, scores were assigned for the above categories. By means of the qualitative criteria which had been decided upon, the other Rorschach scores were determined.

The following procedure was adopted.

1. The examiner read each response on a Rorschach record, and determined whether the whole or a part of the blot had been selected for interpretation. The symbols indicated next to a response or the tracings when used showed this.
2. If it was a response to the whole blot, a score of *W* was given

---

[11] Bound copies of these tables, in mimeographed form, are now available and may be purchased at cost from the Brush Foundation, Western Reserve University, Cleveland, Ohio.

3 If a part had been selected, the examiner investigated to see which detail.

4 The examiner then located the appropriate List to the Card to which this response belonged, by consulting the top sheet for each Card. For example, the answer *"a bat"* for Card I, was a Whole response to be found in List A. *"A candy-jar"* for space-detail (2) Card II was located in List E, Card II.

5. The examiner next determined upon the proper category to which the response belonged —human, animal, anatomical, natural or artificial. For example, *"a bat"* would be found in the list of responses containing animal forms; *"candy-jar"* in the list of artificial forms.

6. The examiner then looked in that section for the response in its proper alphabetical place, *"bat"* under the "*B*'s" and *"candy-jar"* under the "*C*'s"

It should be reiterated that the Frequency Tables were used to determine only the scores, *W*, *D*, *Dr*, *Do*, *DS*, *F*+, *F*—, *O*+, *O*—, and *P* To score for other Rorschach factors, it was necessary to study the record of the subject with the notes and comments of the examiner and to apply the qualitative criteria which had been determined upon It should be emphasized that most of the other categories [*M*, *FC*, *CF*, and *F*(*C*)] include form. For example, *"clowns dancing"* for Card II is movement, yet form is certainly involved It was necessary to consult the tables, therefore, to determine whether the form in the other categories is + or —. Thus the tables were utilized in determining the quality of the form included in other categories.

Thus the entire set of test records were scored. The Frequency Tables as well as the individual tables summarizing the normal details and the popular responses were also utilized. It was found that after experience, the tables could be used easily and with reasonable speed. Scoring could be accomplished with preciseness and objectivity. It was likewise observed that after scoring about 50 records, it was not necessary to consult the lists for every response. The examiner learned the normal details and the popular responses for each test card. Most of the usual responses were likewise remembered.

VI. STUDY OF THE RELIABILITY OF THE SCORING METHOD

Early in the study, when the Frequency Tables were in the course of preparation, the examiner and a second person instructed in the scoring system, scored the first 100 test records for the general Rorschach factors, namely the whole, detail, space detail, form, movement, and color answers without qualifications This scoring was based on references to the original manual of Rorschach, to Rorschach and Oberholzer (26) and to other investigators

The record blank was folded lengthwise at the right, making a flap one and one half inches in width, so that the column in which the score by the first examiner appeared could not be seen. The second judge used this flap, thus scoring without seeing the records made by the first judge.

Comparison of the two sets of scores for these general categories showed a correspondence of 86 per cent.

After the scoring method had been refined, the Frequency Tables compiled and all papers scored by the writer, 100 test records were again selected from the group of 300, and rescored by still another judge who had been instructed in the use of the Frequency Tables and of the qualitative criteria determined upon When the papers were given to this judge, the right and left margins had been folded down so that she was unable to see any of the previous scoring.

Comparison of the two sets of scores by the writer and by the third judge on 100 papers showed a correspondence of 93 per cent.

The last judge reported that she readily learnt to use the Frequency Tables and that responses were easily located. She found it practical to memorize the table of Normal Details and the table of Popular Responses. She likewise reported that the scoring for most of the usual responses was learned and that it was not necessary to use the Tables for each response which was scored

VII. THE PSYCHOGRAM

Results for each subject were summarized on special sheets provided for that purpose This record represented the "psychogram" of the individual. Table 9 is a copy of one of the summary records

Responses were summarized, (*R*) representing the total number of answers given by a subject. The total number of "items" or forms which had been specially mentioned was likewise summarized.

Both the number and percentage of *W*, *D*, *Dr*, *Do*, and *DS* were obtained.

The number of good forms and the percentage of good forms

TABLE 9

SCORING SHEET

Serial Number—3649		*Rorschach Psychogram*
Name—*Brown, John*		Group P. II.
Age—13		Date, Nov 30, 1931
No of Responses—40	Items—64	R T 7" (1st Response)
No of Failures—0		
Mode of Response	*Quality of Response*	*Content*
W 6 % 20	F 25	A 16
dW 2 %	F+ 24	Ad 2
D 27 % 67	F—1	18 % A 45
Dr 1 % 3	%F+ 96%	H 6
Do 0 %	M 9 (7+, 2—)%	Hd 6
DS 4 % 10	Md %	12 % Hd 30
Apper. Mode W—*D*	C	Anat
(Dr) DS	CF 2—	
Succ regular	FC 4+	
	Sum Color 4	Obj. 3 % 8
	F(c) 0%	Arch
	Color Shock No	Bot 2 % 6
	35 F+(R) 87%	Geog
Intelligence		Geol
W 8 or 20%		Fire 1 % 3
%F+ 96% (87%)		Mt
M 9		St
C 4		Sc 4 % 10
%A 45%		Wr
%O+ 20%	P 7 % 18	
Apper. Mode W—*D*	O 10 % 25	
Succession reg.	O+ 8 % 20	
	O—2 % 5	
Erlebnistypus		
M C 9 · 4		
Personality Data.		
I.Q 133 Otis		
I Q. 117 Cleveland		
M.A. 14—1 Porteus		
Ascendency + Submission Test	Score 0	Submissive
Form Board Score 94		
Woodworth-Matthews Inventory	Total Score 94	V. Stable
Discriminative Scores A 10; B 12; C 2; D 3, E 8; F 2		

were computed, this latter according to the Rorschach formula

$$\%F+=\frac{100\ F+}{(F+)+(F-)}.$$

Some investigators such as Oeser (23) prefer to use the formula $\frac{100\ (F+)}{R}$ as a measure of the percentage of clearly perceived forms. The Brush examiners adhered to the original Rorschach formula in the early project, but today both percentages are calculated.

The number of movement answers were summarized.

The color scores were likewise added. Color scores were weighted according to Rorschach's instructions: $C = 1\frac{1}{2}$, $CF = 1$, $FC = \frac{1}{2}$. The sum Color score was then computed.

The content categories were totalled and the incidence of certain classes of forms computed as the percentage of animal forms, human forms, and anatomical forms.

Finally, the number and percentage of $O+$, $O-$, O and P were computed according to the formulae

$$\%O = \frac{100\ (O+ \ +\ O-)}{R}$$

$$\%O+ = \frac{100\ O+}{R}$$

$$\%O- = \frac{100\ O-}{R}$$

$$\%P = \frac{100\ P}{R}$$

The customary practice was to indicate 20% ± or 20% ∓ according to which predominates, good original or poor original answers. The separate percentages were computed in this project with the number and percentage of popular forms.

Finally M was compared with Sum C and the "Erlebnistypus" determined upon.

CONCLUSION

The Brush Foundation initiated a project to determine the reliability and validity of the Rorschach ink-blot test. The first problem embraced was to standardize the technique of the test. For that purpose, 300 adolescent students of the Patrick Henry Junior High School of Cleveland were selected. They were chosen so as to obtain satisfactory distribution in point of chronological and mental ages.

The administration of the Rorschach Ink-blot test was modified so as to increase the uniformity in procedure and the objectivity in scoring. Quantitative and qualitative criteria were determined upon for scoring the various test factors. Frequency Tables were compiled to provide standards of normality for certain of the test categories

Standardization of the procedure and of the technique simplified the administration of the test and increased the efficiency of the examiners It gave greater uniformity and objectivity to the scoring, which later aided materially in determining whether the test is reliable or valid.

The Frequency Tables and the criteria adopted in scoring may prove suggestive to other Rorschach examiners They may be of value in scoring other Rorschach records. Strictly speaking, of course, they apply only to similar groups of subjects, similarly selected. Until further norms are available, however, they may be used tentatively for other Rorschach material, keeping in mind the number and the composition of the sample upon which they are based.

References

1 Beck, S. J The Rorschach tests as applied to a feeble-minded group. *Arch of Psychol*, 1932, No. 136. Pp. 84

2 ———. Configurational tendencies in Rorschach responses *Amer. J Psychol*, 1933, **45**, 433-443.

3 ——— Problems of further research in the Rorschach test. *Amer J Orthopsychiat.* 1935, **5**, 100-115.

4 Behn-Eschenburg, H. Psychische Schuleruntersuchungen mit dem Formdeutversuch. Bern. Bircher, 1921 Pp 66.

5 Binder, H Die Helldunkeldeutungen im psychodiagnostischen Experiment von Rorschach *Schweiz. Arch. Neur.*, 1932-1933, **30**, 1-67, 233-286. Republished in monograph; Orell Fussli: Zurich, 1932 Pp. 123

6 Furrer, A. Über die Bedeutung der "B" im Rorschachshen Versuch. *Imago*, 1925, **11**, 362-365.

7. Ganz, E, & Loosli-Usteri, M Le test de Rorschach applique à 43 garçons anormaux *Arch. de Psychol*, 1934, **24**, 245-255.

8 Gardner, G E. Rorschach test replies and results in 100 normal adults of average *IQ* *Amer. J Orthopsychiat*, 1936, 6, 32-60

9 Guirdham, A. The Rorschach test in epileptics. *J. Ment. Sci*, 1935, **81**, 870-893

10 Hertz, M R. The Rorschach ink-blot test. Unpublished thesis. 1931 Western Reserve University Library, Cleveland, Ohio.

11. ———. The Rorschach ink-blot test: historical summary. *Psychol. Bull*, 1935, **32**, 33-66.

12. ———. Rorschach norms for adolescent age group. *Child Devel.*, 1935, 6, 69-76.

JUARROS, C, & SORIANO, M Le psychodiagnostic de Rorschach chez les enfants anormaux *Comptes Rendus de la IV Conf Int. de Psychotechn*, 595-605 Paris Alcan, 1929

KERR, M. The Rorschach test applied to children *Brit J Psychol*, 1934, **25**, 170-185

LINARES M A Experiments on normal Spanish children with the Rorschach test *Arch neurobiol* 1932, **12**, 693-738

LOOSLI-USTERI, M Le test de Rorschach appliqué à différents groupes d'enfants de 10-13 ans. *Arch de Psychol.*, 1929, **22**, 51-106

LOPFE, A. Uber Rorschachische Formdeutversuche mit 10-13 jahrigen Knaben. *Zsch. f angew Psychol.*, 1925, **26**, 202-253.

MANDOWSKY, C Uber die Bedeutung des Rorschach'schen Formdeutversuchs. *Die Biologie der Person.* (ed T. Brugsch and F H Lewy) 1931, **2**, 1014-1059

MELTZER, H. Personality differences among stuttering children as indicated by the Rorschach test *Amer J Orthopsychiat*, 1934, **4**, 262-282

MONNIER, M Le test psychologique de Rorschach *Encéphale*, 1934, **29**, 189-201, 247-270

MÜLLER, M Der Rorschachsche Formdeutversuch, Seine Schwierigkeiten und Ergebnisse *Ztsch f d ges. Neur und Psychiat.*, 1929, **118**, 599-620

MUNZ, E Die Reaktion des Pyknickes im Rorschachschen psychodiagnostischen Versuch *Ztsch f d. ges Neur. und Psychiat*, 1924, **91**, 26-92.

OESER, O. A Some experiments on the abstraction of form and colour. Part II. Rorschach tests *Brit J Psychol*, 1932, **22**, 287-323.

PFAHLER, G. System der Typenlehren Grundlegung einer padagogischen Typenlehre *Ztsch f Psychol*, 1929, **15**, 334

RORSCHACH, H Psychodiagnostik: Methodik und Ergebnisse eines wahrnehmungsdiagnostischen Experiments Bern Bircher, 1921. Pp. 174.

RORSCHACH, H, & OBERHOLZER, E Zur Auswertung des Formdeutversuchs fur die Psychoanalyse *Ztsch f. d ges Neur. und Psychiat*, 1923, **82**, 240-274 Translated Application of the interpretation of form to psychoanalysis. *J. Nerv. and Ment Dis.*, 1924, **60**, 225-248, 359-379.

RORSCHACH, H Psychodiagnostik. Berlin Huber, 1932. Pp. 227 (Contains references 25 and 26)

SALAS, J El psicodiagnóstico de Rorschach *Arch. de neurobiol*, 1932, **12**, 316-339

———. La classificación de las repuestas en el psicodiagnostico de Rorschach *Arch de neurobiol*, 1933, **8**, 45-80

VERNON, P. E The Rorschach ink-blot test. *Brit J Med Psychol*, 1933, **13**, 89-118; 179-200, 271-291

——— Recent work on the Rorschach test *J. Ment Sci*, 1935, **81**, 894-920.

WELLS, F. L. Personality traits Appendix D *Conference on Individual Psychological Differences.* Dec, 1930 Washington National Research Council, 1931

33 DIMMICK, G. B An application of the Rorschach ink-blot test to three clinical types of Dementia Praecox. *J. of Psychol.*, 1935-36, 1, 61-74

34. KLOPFER, B., & SENDER, S A system of refined scoring symbols *Rorschach Res Exch.*, 1936-37, 1, 19-22

35. SENDER, S., & KLOPFER, B. Application of the Rorschach test to child behavior problems as facilitated by a refinement of the scoring method *Rorschach Res. Exch.*, 1936, 1, 5-17.

36. THORNTON, G R. A note on the scoring of movement in the Rorschach test. *Amer. J. Psychol.*, 1936, 48, 524-525.

Developmental Health Inquiry
Western Reserve Medical School
Cleveland, Ohio

The Journal of Genetic Psychology, 1938, 52, 65-77

THE BODY JERK OF THE NEONATE*

Department of Psychology, The Ohio State University

ISABELLE F WAGNER

CONSIDERATION OF PREVIOUS CONCEPTS

It is characteristic in the study of infant behavior to give names to casually observed reaction patterns and thereafter to use those names as if they referred to distinct, indisputable forms of behavior Often the original name, innocently enough, is used merely for convenience in referring to such casually observed behavior; but subsequent investigators seize upon the term and read into it a specific, well defined connotation which is henceforth attributed unjustly to the originator of the term Consequently later investigators who have the temerity to examine more closely the behavior patterns thus named, usually discover that the behavior in question is not so clear-cut and invariable as previously assumed

This is especially true in the case of the body jerk, although superficially no difficulty is apparent. In this paper we shall examine some of the earlier concepts and determine from an analysis of a large group of data whether or not the specific patterns described actually exist.

Dennis (5), after a careful review of the literature describing or classifying the responses of the newborn infant, appeared to find a clear differentiation between the "startle" and the "bodily jerk," each of which he described specifically. The "startle" refers to the "surprise reaction" of Preyer, the "*Umklammerungsreflex*" of Moro, the "fear reaction" of Watson, the "*Schreckreaction*" of Cemach, Buhler, Peiper, and others, and the "starting" of Bryan The arms fly apart, fingers spread, legs extend, and the head is thrown backward In a minor degree of the startle, only the arms may be affected. Whether crying follows seems to depend on the intensity and duration of the stimulus. The startle may be elicited by loud noises and falling (Watson and Morgan), it may be a vestibular response (Moro, Magnus, DeKleijn); it may be observed when the infant is going to sleep (Bryan). It may result from touch stimuli (Preyer); from stretching hip and knee, moving other bodily parts, or stimulating large areas of the body with hot or cold stimuli (Freudenberg); from a blow on the chest or abdomen (Freudenberg and Peiper); or from strong stimulation of any receptor (Peiper).

---

*Recommended for publication by F C Dockeray, and received in the Editorial Office on April 20, 1937

In the "bodily jerk," on the other hand, the arms and legs flex strongly and jerk upward in response to loud noise. Here are included Champneys' "spasmodic start" and Irwin's "body jerk."

But it is possible that the apparent simplicity of Dennis' classification is largely superficial, and that the terms borrowed from other investigators were not considered in the light of their original meanings An examination of a few of the individual concepts included in this bimodal classification indicates that the above differentiation is open to question.

Preyer (7) gave the name "fright" to the reaction to sound, and sometimes to touch, and described it as: movement of the eyelids, starting, throwing up the arms, and screaming; various degrees of this reaction are: quiver of eyelids; twisting of head; starting, accompanied by violent quivering of head, arms, and upper part of body

Blanton (1) found that the response to dropping consisted of an upward movement of the arms, and sometimes holding of the breath, while in response to a sound the infant threw the arms forward, moving the entire body. She also described such responses as a "jump" and as "convulsive movements," but did not specify the leg movements involved.

Watson (9) later described the response to noise or loss of support as follows: catching of breath, clutching randomly with hands, sudden closure of eyelids, puckering of lips, and crying. Nothing is said here about the leg movements.

Bryan (2) did not describe the "starting" to which she referred, just as Champneys (4) had not described the "spasmodic start" which he declared resulted from any jar or vibration. Bryan found the reaction occurred to certain sounds and also in the process of going to sleep.

Buhler (3) observed that sudden or strong sensory stimuli, as well as rough changes in position, produced a convulsive quivering or "contraction" followed by crying This "shock reaction" often included the utterance of a sound of fright Other types of "negative expressional behavior" or "fright" were: (*a*) crooking of body, frowning of forehead, straining of legs, making fists, firm pressing together of eyelids, (*b*) throwing back of head, lifting up of body, stretching out of arms, spreading of fingers, opening the eyes wide It is not clear here that Buhler was referring merely to the "startle" described by Dennis, the "shock reaction" could well include the "bodily jerk"

In Irwin's (6) analysis of the activities of newborn infants, he listed the "quick body jerk" but no "startle," for apparently "body jerk" was used as a more general term including the other as one of its forms In the "body jerk" there is a sudden tensing of trunk muscles, hips and legs are quickly flexed, likewise arms are flexed and hugged to the chest. *Often the opposite reaction occurs*—arms and legs may be suddenly extended. All segments of the infant participate in the reaction. The shudder, gasp, back arching, stretching, or vocal concomitants are here listed separately, not being considered essential parts of the reaction described.

What we have, therefore, is in reality a variety of responses to a variety of stimuli, where terms such as "start" or "jerk" are frequently used. All agree as to the suddenness and jerkiness of the responses discussed, but there is little agreement as to details.

RESULTS OF THE PRESENT EXPERIMENT

To decide whether or not there are distinct responses such as a "startle" and "body jerk" and to discover their essential characteristics, we have recorded all the "jerky" responses made by 197 newborn infants to over 5000 separate presentations of a variety of tactual, pain, olfactory, and auditory stimuli For 97 of these subjects the jerky responses were described specifically The details of the experimental procedure, subjects, and apparatus are given in an earlier article (8). It is sufficient to note here that records consisted of a pneumographic curve, stimulus, and running annotations concerning observed behavior, all recorded on a moving polygraph tape.

Those responses which were finally tabulated for the present study included all reactions involving a jerk or sudden tensing of the trunk as well as some form of limb movement These reactions were listed with reference to whether they were elicited by one of the experimental stimuli or by a stimulus not apparent to the experimenter—that is, with reference to their occurrence immediately following the presentation of a specific stimulus, or somewhere between successive presentations. However, in the succeeding analysis these two groups of jerky reactions were combined, since careful inspection revealed no consistent differences between the two sets. In other words, so-called "spontaneous" jerky movements are not qualitatively different from those elicited by stimuli externally applied.

First of all, the 368 jerks which were described in specific detail were examined to determine whether or not they could be grouped into specific patterns.

TABLE 1

PER CENT OF OCCURENCE OF VARIOUS PATTERNS OF JERKY REACTIONS

Pattern	*Number*	*Per Cent*
"Startle"	65	17.7
"Bodily Jerk"	65	17.7
Opposite	51	13.9
2 Arms, 1 Leg	32	8.7
2 Legs, 1 Arm	8	2.2
Arm Only	15	4.1
Leg Only	7	1.9

To ascertain the number of "startles," as described by Dennis, we included the following responses, regardless of any other components involved: arms and legs extended; arms extended and legs raised; arms extended without leg movement. Of the 368 responses only 17.7 per cent fit into this classification; of these, 1.1 per cent are arm extensions without leg movements.

In Dennis' "bodily jerk" classification were included all responses involving flexion of arms and legs, again regardless of other pattern components. By coincidence, we find the same per cent—17.7—as of "startles."

But these two classifications, loose though they are, include only a third of the data. There are 51 cases—13.9 per cent—where leg and arm movements are opposite; that is, legs extend while arms flex, or vice versa There are some asymetrical patterns as well; there are eight instances of movement of both legs and one arm (2.2%), if we make no distinction between flexion and extension, and 32 instances (8.7%) where both arms and only one leg move. In 15 cases (4.1%) arm movement occurs without leg movement;

this does not include the four cases classed as "startles," but does include one case of fingers only, one of right arm only, and one of left arm only. In 7 instances (1.9%) leg movement occurs without arm movement.

Even though we consider the foregoing responses adequately classified, we still have 33.8 per cent remaining for which it is difficult to devise convenient categories. Examples are: slight jerk arms and legs, quick start of arms, with flexion of right leg and extension of left leg, quick twist and inspiration, with right hand adducted over face, extension and stiffening of legs, followed by flexion of legs and adduction of left arm; tightening of eyelids, jerk of head to right, and slight stir of arms and legs, quick flexion of legs, right arm up vertically over face, and left arm clutched to body.

Even the foregoing classifications have inherent difficulties which render the groupings rather artificial. If we looked for examples concurrent in every detail with the descriptions given by Dennis, we should have only one or two of each! To get a semblance of agreement, we must discard such factors as vocal concomitants, deep inspirations or gasps, head movements, and eyelid, mouth, and other facial movements. Furthermore, we must omit the finger-spreading factor, which occurs in less than half (30) of the "startles," where arms extended or abducted; it may even occur (16 cases) where the arms are flexed or adducted.

Moreover, it is often hard to distinguish flexion from extension, and abduction from adduction. Only those who have attempted to describe the infant's limb movement can realize the difficulties involved in verbalizing these three-dimensional movements. If vertical movement—lifting or raising of limbs—is also included, even fewer cases would be cited, for arms and legs may "go apart" (abduct) or "flex strongly" without being "jerked upward."

The differentiation would become even more involved if the intensity or vigor of the responses were considered. Most of the investigators using a term like "startle" or "body jerk" indicated a vigorous response. But where can the line be drawn? In our data there are responses of all degrees of intensity, varying from slight jerks to convulsive body movements. In some instances a given response may readily be said to be more vigorous than another, but it would be less simple to establish an absolute standard of vigor for distinguishing a "body jerk" or "startle."

If absolute vigor were considered, then the relative vigor of the limb movements themselves should be discussed. Often one set of limbs moves more vigorously than the other, just as one arm or one leg may make a more jerky or more extensive movement than its bilateral mate.

It may be objected here that the more vigorous responses are less variable in pattern, hence if we had used stronger types of stimulation, the data secured would have revealed less variability. This objection is overruled by the fact that even the most vigorous responses obtained in this study varied greatly in pattern.

Another complication would arise if we considered the specific sequence of movements involved in the responses. While in many cases the jerk of trunk and limbs occurs simultaneously, in others the trunk jerk precedes the limb movement. Sometimes the arm and leg movements, while in close temporal sequence, are not simultaneous, or one limb may precede the one opposite. Still another type of sequence may be obtained: In some cases, flexion of limbs is immediately followed by extension, or extension by flexion, although we could ascertain no relationship between this type of sequence and any other factor. Some infants, moreover, whatever the specific movements involved in the body jerk, immediately resume their former position, while others, though they also relax, retain certain modifications of position occasioned by the body jerk.

The "bodily jerk" and the "startle" as defined by Dennis cannot be differentiated from each other or from other responses in terms of the type of stimuli required to elicit them. Dennis' summary indicates the variety of stimuli assigned to the "startle" response, though he gives only loud sounds as the stimuli for the body jerk. Inspection of our own data reveals that those loosely classified as "startles" or "bodily jerks" can at times be elicited by any one of the stimuli used in this experiment—tactual, pain, olfactory, or auditory. Furthermore, there is no consistent dominance of parts in terms of vigor of response with reference to the area stimulated. What frequently happens is that a local movement occurs, such as a foot flexion in response to a toe prick, and this is immediately followed by a jerky body movement which may be partially, if not completely, initiated by kinesthetic stimuli afforded by the local movement. Hence the differentiation in terms of stimulus can at times be made with respect to an initial local response, but not with respect to the jerky response itself.

Summarizing our discussion up to this point, we find that: (*a*) Only by very loose descriptions of the "startle" and "body jerk" can we find an appreciable number of examples of each. (*b*) These descriptions are only two instances of a variety of patterns which may be adopted for classification. (*c*) Fully half of the data cannot be classified conveniently in terms of patterns of limb flexion and extension, abduction and adduction. (*d*) The most noticeable, as well as the most neglected, factor is the high degree of variability and complexity in responses involving a sudden trunk jerk and limb movement

A REVISION OF PREVIOUS CONCEPTS

Having recognized the inadequacy of the concepts of "startle" and "bodily jerk" heretofore suggested, we are left with one of two alternatives: We may abandon the terms entirely in our discussion of infant behavior, or we may adopt a broader definition of the term "body jerk," recognizing the wide variations of response involved. Logically, of course, we could use the term "startle" instead of "body jerk" if we retain only one of the two terms However, "body jerk" seems preferable because it is more objectively descriptive of the behavior indicated, and does not convey implications, as does the term "startle," of an interpretation in terms of adult behavior.

Of the two alternatives, a revised definition of "body jerk" appears more desirable. Our data are indicative of the frequent occurrence of sudden, jerky body movements as contrasted with slower responses of greater duration. It is convenient to have some general term to refer to the former class, and there can be no objection to selecting the term "body jerk" if we keep in mind its revised connotation

The most suitable definition of "body jerk," in terms broad enough to be useful here, appears to be that adopted in the collection of the data for this investigation: any sudden jerk or tensing of the trunk, plus limb movement.

ANALYSIS OF THE BODY JERK

In the light of this definition, we can proceed to point out certain trends in the data other than those already mentioned.

TABLE 2

PER CENT OF OCCURRENCE OF VARIOUS FACTORS IN BODY JERK

(368 body jerks—97 infants)

Factor	*Per Cent*
Vocal	20.7
Deep Inspiration	64.4
Eyelid	6.8
Mouth	6.0
Facial	2.7
Head	16.6

To answer the question as to whether or not such factors as vocal concomitants, eyelid movement, and the like are essential parts of the body jerk, we can examine the data to find the per cent of cases in which these factors occur. In 20.7 per cent of the 368 cases given in detail, vocal sounds occur—audible inspiration or gasp, audible expiration or grunt, throat sound, whimper, or cry. The majority of these are audible inspirations; only 17 are audible expirations, 2 whimpers, 4 throat sounds, and 6 cries. In only 2 cases does crying follow the body jerk.

Specific eyelid movement occurs in 25 cases (6.8%), and specific mouth movement about as frequently—6.0 per cent. Other facial movements such as those known as "frown" and "grimace," are infrequent, occurring in only 10 cases (2.7%). In a larger number of cases—61 (16.6%)—specific head movement occurs, such as head arched back, head rolled to one side, or slight head stir.

Nor is the deep inspiration, audible or inaudible, an essential component of the body jerk. Some investigators have considered the "gasp" or audible inspiration characteristic of the pattern known as "startle," but this is not true even of those data which roughly fit this term. The audible inspiration occurs in only 12.4 per cent of our 368 cases. A deep inspiration, whether audible or inaudible, occurs in about two-thirds of the cases—64.4 per cent. It is often difficult to distinguish a deep inspiration from the effect of body motility on the breathing curve, but a deep inspiration was usually recorded wherever there was a sudden sharp rise in this curve. Sometimes the inspiration and body jerk occur simultaneously; occasionally the deep inspiration immediately follows the body jerk, in turn followed by a general stir, in a few cases the deep inspiration precedes the jerking of the body.

The lack of a close relationship between the deep inspiration and the body jerk is borne out by still other facts. In many instances where the infant is very quiet and shows no facial movement of any sort, periodic deep inspirations without body movement may occur just as periodic "spontaneous" body jerks occur in others. The deep inspiration is not necessarily the antecedent of the body jerk, since it can thus occur alone and since many of the "spontaneous" body jerks just mentioned give no evidence of a sudden inspiration. Furthermore, the deep inspiration may occur in many types of general activity where the segmental movements are too slow in their sequence to be classed as a body jerk.

It should be noted also that there is no clear-cut line of demarcation between the body jerk and the general body stir. The most obvious difference between the two is the speed of the response. Wherever a series of movements is relatively slow rather than sudden and jerky, that pattern of general activity is called a "general stir." Usually the limb movements in a body jerk are symmetrical, while those in a "stir" are not. But there are always borderline cases in which speed and symmetry of movement are not sufficient criteria, because behavior categories are always arbitrary; behavior is a continuum and not a mosaic of distinct reaction patterns.

Leaving this descriptive analysis of the body jerk, we may next examine a larger body of data secured from the total of 197 infants, since, for the first 100 of these, body jerks were recorded but not described in detail. These body jerks, 510 in number, are responses to specific stimuli, and include all the data previously described except the 150 body jerks for which there was no apparent stimulus.

The question arises as to whether, within the first 10 days of life, there is any variation, with increasing age, in the number of body jerks elicited. The per cent of total stimulations eliciting a body-jerk response is given for each age group.

Age in Days[1]	Per Cent
1	9.5
2	17.1
3	13.9
4	8.7
5	3.8
6	8.3
7	8.3
8	9.5
9	7.2
10	10.8

Body jerks as responses to externally applied stimuli seem most frequent during the first few days, especially during the second day, thereafter decreasing to a minimum at five days and rising somewhat in frequency toward the end of the ten-day period.

A similar trend is shown for the 150 body jerks occurring without an apparent stimulus:

Age in days	Per cent
0	13.3
1	20.0
2	12.7
3	10.7
4	4.0
5	5.3
6	2.0
7	2.7
8	17.0
9	4.7
10	6.7

Here the maximal frequency occurs during the first day, and the minimal during the sixth, with an irregular rise in subsequent age groups.

Another question arises as to whether body jerks in the newborn infant as in the adult characterize a period of apparent going to sleep. In a previous article (8) the author determined a series of stages of depth of sleep in terms of overt motility. These stages, ranked from least to greatest motility, or from deepest to lightest sleep, were:

A_{1_a}
A_{1_b}
B_1
A_2, A_3, B_2[1]
B_3
C_2, C_3
$C_6+B_4+B_5$

---

3 days—72 to 96 hours; etc. For part of the data a 0-day classification (birth to 24 hours) was used, but since for the remaining data only the more inclusive 1-day group was used, the latter form was adopted for convenience in this particular table.

[1] Stages A_2, A_3, and B_2 were grouped together because, while not significantly different from one another, each was significantly different from the stage preceding and the stage following. The same holds true for C_2 and C_3. Stages C_6, B_4, and B_5 were significantly different from C_2 and C_3 only when combined.

Specific description of these stages may be found in the earlier article.

Comparison may be made of the per cents of body jerks and general stirs occurring in each condition:

Stage	Body Jerks	Stirs
A_{1_a}	10.4	4.7
A_{1_b}	12.5	7.3
B_1	15.5	9.3
A_2, A_3, B_2	10.1, 4.0, 8.3	20.6, 16.9, 20.9
B_3	5.5	25.9
C_1, C_2	11.2, 10.2	36.6, 34.7
$C_3+B_4+B_5$	6.6	55.9

It appears that general-stir responses occur with steadily increasing frequency as the infant becomes more active, or less deeply asleep, while body jerks decrease somewhat. Body-jerk responses are most frequent in stage B_1, when the infant is beginning to move body members occasionally, but as yet evinces no eyelid or mouth movement.

The sequence of decreasing body jerks is more regular if we distinguish between (1) body jerks after which the infant quiets immediately, and (2) body jerks followed immediately by general stirs. The second form seems to be an intermediate stage between the body jerk and the general stir. The per cents are then as follows.

Stage	Body Jerk (1)	Body Jerk (2)
A_{1_a}	9.2	1.2
A_{1_b}	11.5	1.0
B_1	11.9	3.5
A_2, A_3, B_2	7.7, 1.6, 6.6	2.4, 2.4, 1.7
B_3	4.7	0.8
C_1, C_2	5.9, 6.0	5.3, 4.2
$C_3+B_4+B_5$	4.4	2.2

The high percentages for C_2 and C_1 in the table where no such distinction is made, are thus shown to be due to the large proportion of responses in which the body jerk was protracted in a general stir.

Referring again to the 150 "spontaneous" body jerks, we find a similar trend. The per cent occurring in each condition is:

Stage	Body Jerks	
A_{1_a}	26.7	
A_{1_b}	34.7	
B_1	6.7	
A_2, A_3, B_2	6.0, 2.0, 11.3	Mean—6.4
B_3	4.7	
C_2, C_1	4.7, 1.3	Mean—3.0
$C_5+B_4+B_5$	2.0	

The largest number of "spontaneous" body jerks thus appear in stages A_{1_a} and A_{1_b}, when the infant is generally quiet and shows no eyelid or mouth movement. As the infant becomes more active, or less deeply asleep, these body jerks decrease steadily, almost disappearing in the most active stages.

SUMMARY

The following conclusions may be drawn from this study:

1. There has been no general agreement heretofore as to the specific behavior pattern indicated by the terms "startle" and "body jerk." At best the patterns referred to were highly specific forms of a wide variety of responses involving the jerking of the body.

2. It is preferable to use "body jerk" as a general term referring to any response involving a trunk jerk and limb movement. The patterns previously cited by investigators thus become only special types of body jerk.

3. Concerning the body jerk as we now define it, the following facts have been ascertained:

a. The variability of the body jerk is its most striking characteristic. The largest proportion of the body jerks cannot be classified in terms of any well defined patterns.

b. Specific vocal, head, eyelid, mouth, and other facial components are not essential factors in the body jerk. Their relative frequency of occurrence, from highest to lowest, is in the order named.

c. The deep inspiration, though apparently not essential to the body jerk, occurs in about two-thirds of the cases.

d. There is no distinct line of cleavage between the body-jerk and general-stir categories, though they usually differ in speed and symmetry of movement.

e. Body jerks, whether or not responses to overt stimuli, seem most frequent during the first few days, thereafter decreasing to a minimum at the middle of the ten-day period, and rising irregularly in frequency toward the end.

f. Body jerks as a rule decrease in frequency as the infant becomes less deeply asleep, while general stirs increase

REFERENCES

1. BLANTON, M. G. Behavior of the human infant during the first thirty days of life. *Psychol. Rev.*, 1917, 24, 456-483.
2. BRYAN, E. S. Variations in the responses of infants in the first ten days of post-natal life. *Child Devel.*, 1930, 1, 56-77.
3. BÜHLER, C. The first year of life. New York: John Day, 1930.
4. CHAMPNEYS, F. H. Notes on an infant. *Mind*, 1881, 6, 104-107.
5. DENNIS, W. A description and classification of the responses of the newborn infant. *Psychol. Bull.*, 1934, 31, 5-22.
6. IRWIN, O. C. The amount and nature of activities of newborn infants under constant external stimulating conditions during the first ten days of life. *Genet. Psychol. Monog.*, 1930, 8, 1-92.
7. PREYER, W. The mind of the child. New York: Appleton, 1914.
8. WAGNER, I. F. The establishment of a criterion of depth of sleep in the newborn infant. *J. Genet. Psychol.*, 1937, 51, 17-59.
9. WATSON, J. B. Psychology from the standpoint of a behaviorist. Philadelphia: Lippincott, 1924.

Department of Psychology
The Ohio State University
Columbus, Ohio

The Journal of Genetic Psychology, 1938, 52, 79-149

A STUDY OF PERSONALITY IN YOUNG CHILDREN BY MEANS OF A SERIES OF RATING SCALES*

The Merrill-Palmer School

Katherine Elliott Roberts and Rachel Stutsman Ball

Introduction

Observation of infants and young children reveals that individual characteristics begin to emerge quite early in life. Many investigators have shown that some aspects of behavior follow a fairly fixed sequence of development. Thus, it is characteristic of a baby, first, to pay no attention to strangers, then to recognize them, and finally to cry in their presence And so with some other aspects of behavior in infancy and childhood But over and above these expected sequences of development there are individual ways of behaving. Two children entering nursery school may show extremely different reactions. One is immediately at ease, fits into the day's routine without any trouble, is friendly with other children; in short, adjusts well to the new situation. The other is quiet, withdraws from other children, has difficulty in eating and sleeping, and in general has to make a real effort to learn to fit into the regime. These different ways of behaving may or may not be fundamental indexes of the child's personality, but from the point of view of observable behavior they serve to distinguish one individual from another

In the Merrill-Palmer nursery school records of the children's behavior have been kept since the early days of the School One kind of record is called the "Daily Diary." At regular intervals the teacher dictates reports of significant events and behavior in the nursery school which are placed in the files of the children concerned Much valuable information has been recorded in this manner, and the method has many advantages, but there are also certain drawbacks, i e , it is extremely time-consuming; significant items are not always followed from one time period to another, the material is purely descriptive and therefore often difficult to evaluate; the information recorded is not always comparable for different children; spectacular events are likely to be recorded at the expense of the

---

*Received in the Editorial Office on April 20, 1937

more usual daily behavior of the child; and more information is recorded for some children than for others. Several other types of records, such as observations and anecdotes, have also been used to describe the behavior of the children

However, as the years went by and more and more children entered the nursery school it became apparent that some more effective method of studying and recording personality was needed. As a result Dr. Rachel Stutsman Ball and Miss Leone Chesire began in 1929 to study personality by means of rating scales. Their purpose was (*a*) to construct scales for rating various aspects of personality and (*b*) to evaluate these scales in terms of the agreement of raters, the variability of the ratings over a period of years, and the adequacy of the scales for describing the development of personality.

With the exception of one or two studies in which some of the children rated were enrolled elsewhere, all the children rated were Merrill-Palmer children, and the raters were teachers, students, and parents at the School.

We publish the study at this time, not because it is a finished piece of research, but because after using the scales for seven years we think they may have some value in stimulating further research in the personality development of young children. There are many weaknesses in the scales which will be immediately apparent to the reader. These we are cognizant of and will discuss. At the same time, the scales and the methods employed have been tested more rigorously than is usually the case. Reliability has been tested through correlations, not of various ratings made a single time, but of repeated ratings of many raters over an extended period. Further, a number of the raters were experienced, and all had excellent opportunities to know the children through seeing them daily in a wide variety of situations in the nursery school. The children rated were a fair-sized group, and many of them were rated continuously for several years. The study also supplies data permitting the construction of personality pictures of individual children, as rated from time to time over a period of years. To some degree, the study is an examination of rating technique as well as an evaluation of the nine scales.

THE DEVELOPMENT OF THE SCALES

In 1929-30, as a first step in developing the scales, many sources of descriptive material about children were consulted. Books on child psychology, the "daily diaries" of the nursery-school teachers, check lists of adjectives, student observations, and case studies were some of the sources yielding phrases descriptive of child behavior. Hundreds of such phrases were collected and after careful consideration were sorted into nine major categories: ascendance-submission, sociability with other children, compliance with routine, response to authority, respect for property rights, tendency to face reality, independence of adult affection and attention, attractiveness of personality, and physical attractiveness.

Thurstone's (14) method of measuring attitudes was used as a basis for scaling the items. It must be kept in mind, however, that whereas Thurstone's attitude scales measure an adult's opinion about war, the church, crime, and so on, these personality scales measure a teacher's opinion about a child as judged from his behavior. In other words, the ratings represent an adult's opinion about a child rather than an adult's opinion reflecting his own attitudes. In order to construct the scales, series of statements intended to describe each of the nine traits to be measured were grouped in nine categories. Each of these nine groups of statements was given to from 50 to 70 judges (Merrill-Palmer staff and students) who were asked to arrange them in 11 groups, 1 representing the least degree of the characteristic measured, 6 the middle value, and 11 the greatest possible degree. Judges were not asked to rate specific children but merely to sort statements about these nine aspects of personality of children in general. This process assumes that the statements submitted to the raters can be thought of as being distributed along a linear trait continuum.

On the basis of these results, scale values for each statement were determined graphically by means of Thurstone's *Outline for Calculating the Scale Value and the Ambiguity of an Opinion*. By this method, the scale value of each statement is its median placement on the 1 to 11 scale by the group of 50 to 70 judges. For facility in scoring, each value was multiplied by 10 in order to eliminate the decimal point. Thus the scores, instead of ranging from 0.1 to 11, range from 1 to 110. Statements were discarded from or retained in the final scale on the basis of three criteria: ambiguity, irrelevance, and popularity (4).

The criterion of ambiguity was a very rough estimate of the dispersion of each statement. Except for the Response to Authority scale, no quartiles or standard deviations were calculated. A statement having a seven to ten interval spread was usually discarded as ambiguous unless it had a marked modal tendency. The best statements had a five-interval range with a definite mode in the center interval. A tendency toward a rectangular distribution in two to four adjacent intervals was likewise a criterion of ambiguity. Bimodality, and sometimes trimodality, was a criterion of irrelevance. Popularity refers to the total number of times a statement was checked. This criterion was applied after the scales had been set up and used in trial forms to rate nursery-school children. Statements checked for almost every child were discarded as not discriminating. Those very infrequently used were judged as not applicable either to the nursery-school situation or to the young child.

The scale value distributions were smoothed and corrected for end effect, and corrected scale values were obtained.

Table 1 shows the number of judges for each scale and the number of statements in each of the nine scales.

TABLE 1

Scale	No judges	No statements sorted	No statements in mimeographed form	No statements in final form
Response to authority	65	45	45	28
Respect for property rights	50	45	45	30
Independence of adult affection	70	86	51	33
Sociability	70	100	50	50
Ascendance-submission	70	97	48	48
Attractiveness of personality	70	150	75	75
Tendency to face reality	70	189	89	43
Compliance with routine	69	185	31	31
Physical attractiveness (1)	(70	(86	68	68
Physical attractiveness (2)	(70	(48		

USE OF THE SCALES AT THE MERRILL-PALMER SCHOOL

In order to make the scores on the various scales comparable, percentiles were computed. All ratings used for the percentile tables were made by the nursery-school teachers for the children from two to five years of age, and beyond that age by the leaders of the recreational clubs. Only one rater of the nursery-school children rated them over a period of years. The assistant nursery-school teacher is usually chosen from the student group of the previous year. Thus, she has had the opportunity of observing and working with the children for a year before she is asked to rate the child. This is not true, of course, for the new children. But in this case the assistant teacher has the same opportunity as the head teacher to know the children before she is asked to rate them.

After children "graduate" from the nursery school, they come back once a week during the school year for recreational club meetings. All the ratings used for the percentiles beyond five years were made on these club children. The club leaders have had some acquaintance with these children during their stay in the nursery school and consequently have a good background upon which to base their observations in the clubs.

Since the scales were devised primarily to study nursery-school children, some of the statements are in terms of nursery-school situations. This is especially true of the scale *Compliance with Routine,* and is true to some extent of *Respect for Property Rights* and *Response to Authority*. However, all except *Compliance with Routine* have been used with children beyond nursery-school age. Under the direction of Mrs. Elise Hatt Campbell the recreational-club children were rated by the club leaders. Although percentiles have been reported for these older children, the scales need much more study for the ages beyond six years before they can be considered adequate. The reliability of the scales for older children needs to be established, as well as their adequacy to measure personality development with increasing age The same children should be followed over a period of years and the results of the ratings should be studied in relation to other data which may be available on personality development

When the scales were first used, ratings were made at frequent intervals in order to study the effect of the number and quality of raters As would be expected, it was found that the nursery-school teachers were better raters than students working in the school or staff members who had only a superficial knowledge of the children It is essential that the raters should know the child well in a variety of situations

Since the fall of 1933 ratings of the children have been made by the nursery-school teachers and club leaders twice a year, in the fall and in the spring.

The scales are so constructed that the rater may check only those items which she feels reasonably certain describe the child. She may check one item or many. Ratings of this type make it easy to read off the statements checked, and the combination of statements gives a much more adequate conception of the aspects of behavior included in the judgment than do methods that describe the child merely with a qualitative statement, as for example, *very sociable* or *not at all sociable*. The mean scale value of the various items checked is the child's score for a given rater. Because there is considerable variation among raters, the average of two ratings is used as the child's score on a given scale at any one time. The percentiles for all the scales are based on average scores, the average of two raters who know the child well. We realize that more than two raters would be desirable, but since we have attempted to devise a practical scale for use primarily in nursery schools and since most nursery schools do not have more than two teachers who know the child well enough to rate him, we have based our percentiles on the ratings of two teachers. It is undoubtedly better to use the average of two good ratings than of several poor ones. The average of a dozen inadequate ratings will not give as true a picture of a child's behavior as a single rating made by a well-trained rater who knows the child well. Each scale will be discussed separately.

Sociability with Other Children

Since personality development takes place in a social medium, the aspect of behavior which has been termed in this study "sociability with other children" is of primary importance.

A copy of the sociability scale follows (Table 2). Scale values[1] are given in the column at the left of the statements for the convenience of the reader, but these values of course do not appear on the rating sheets given to the raters.

The 50 items are arranged in random order, the statements with

---

[1] The scale values published in this study are slightly different from the values on the printed scoring keys which have been available for some time from the Merrill-Palmer School. The revised values were made after the distributions were smoothed and corrected for end effect.

TABLE 2

MERRILL-PALMER SCHOOL PERSONALITY RATING

Date—

Name—

Rated by—

Directions for Rating.

Check only those statements which you feel are really true of the child Do not guess if you are not reasonably sure A few true statements are better than many half true ones

SOCIABILITY WITH OTHER CHILDREN

Scale value		
99	( )	Makes friends with other children easily
25	( )	Finds it difficult to approach other children and make friends.
105	( )	Makes friends with any child who happens to be around him.
7	( )	Resents interest shown by other children; wants to be left alone.
15	( )	Does not respond to friendly advances.
56	( )	Has a particular friend (Underline: own sex, opposite sex) whom he admires very much (hero-worship or crush).
47	( )	Tries to make entry into group of children but fails.
20	( )	Other children refuse to play with him
26	( )	Is ridiculed or "picked on" by other children.
101	( )	Unhappy if he is not playing with other children
18	( )	So absorbed in his own ideas that he pays no attention to other *children.*
51	( )	Plays only with a gang or group of specific children; refusing to play with others.
87	( )	Contributes to the ideas of the group though not a leader (co-operative companion).
30	( )	Refuses to cooperate with other children unless he is the leader.
45	( )	Hesitant in making suggestions to other children
83	( )	Interest of other children spurs him on to activity.
82	( )	Assumes a protective attitude toward other children (Underline. same sex, opposite sex)
84	( )	Usually pleasant with other children.
21	( )	Often abrupt and surly with other children
99	( )	Has a pleasant manner of securing cooperation from other children
49	( )	Has strong likes and dislikes for other children
53	( )	Rather placid attitude toward other children; neither likes nor dislikes them to any degree
78	( )	Quarrels with other children only over serious issues
18	( )	Quarrels with other children often over trivial things
81	( )	Seldom quarrels with other children over trivial matters.
33	( )	Picks on one particular child.

TABLE 2 (*continued*)

Scale value		
10	( )	Rough and mean with other children
17	( )	Teases or torments younger children.
39	( )	Hurts other children often due to carelessness.
29	( )	Impatient with other children
26	( )	Enjoys seeing other children reprimanded.
17	( )	Ridicules other children.
23	( )	Very critical of other children.
98	( )	Is a good sport when he loses to some other child
93	( )	Is sympathetic toward other children.
95	( )	Affectionate toward other children
98	( )	Very thoughtful of other children.
90	( )	Tries to help the smaller children
27	( )	Resents aid from other children
95	( )	Forgiving of other children who have hurt him, taken his belongings, etc.
32	( )	Tries to get even with a child with whom he is angry
94	( )	Talks to other children a great deal
17	( )	Seldom talks to other children
35	( )	Cries easily in playing with other children
99	( )	Generous in letting other children share activities and possessions.
9	( )	Selfish with other children, doesn't want to share possessions or let them enter into his activities.
28	( )	Doesn't want other children to get attention from adults.
52	( )	Attention from other children leads him to "show off" or act silly.
34	( )	Jealous if other children play with specific child whom he likes very much
89	( )	Not jealous if other children play with his particular friends

high scale values being indicative of sociability and those with low scale values of lack of sociability. For most of the statements there are opposites, but the paired statements do not necessarily appear successively. For example: "*Talks to other children a great deal*" and "*Seldom talks to other children*"; "*Generous in letting other children share activities and possessions*" and "*Selfish with other children; doesn't want to share possessions or let them enter into activities*"

Variability of the Ratings

An outstanding characteristic of the ratings is their variability. Not only is there considerable difference between the two raters at the same time; there is also a difference between the same rater at different times, and likewise between the two raters' average scores at different times. This variability holds true for all nine scales.

There is no conclusive evidence to indicate definitely whether it is related to the scales themselves, to the raters, or to the fact that the aspects of personality rated are so variable that this method of judging them is not adequate. It is true, however, that the scales compare favorably with other rating scales in reliability. Most of the correlations were between .45 and .85, the majority being .60 or above. According to Symonds (13), the typical reliability for rating scales is .55.

Two Merrill-Palmer students, Miss Mary Jane Strayer (12) and Miss Agatha Bowley (3), made special studies of sociability, using these scales.[2] Some of their findings have been utilized in this report.

TABLE 3

SOCIABILITY WITH OTHER CHILDREN

Form *A*

Scale value	
99	Makes friends with other children easily
7	Resents interest shown by other children, wants to be left alone.
15	Does not respond to friendly advances.
47	Tries to make entry into group of children but fails.
20	*Other children refuse to play with him*
26	Is ridiculed or "picked on" by other children.
30	Refuses to cooperate with other children unless he is the leader.
82	Assumes a protective attitude toward other children (Underline: same sex, opposite sex)
84	Usually pleasant with other children
21	Often abrupt and surly with other children.
49	*Has strong likes and dislikes for other children.*
53	Rather placid attitude toward other children; neither likes nor dislikes them to any degree
18	Quarrels with other children often over trivial things
81	Seldom quarrels with other children over trivial matters.
39	Hurts other children often due to carelessness
17	Ridicules other children
98	*Is a good sport when he loses to some other child*
95	Affectionate toward other children.
98	Very thoughtful of other children
27	Resents aid from other children.
32	Tries to get even with a child with whom he is angry
94	Talks to other children a great deal
34	Jealous if other children play with a specific child whom he likes very much.
89	Not jealous if other children play with his particular friends.

---

[2] These and other student studies referred to were done under the direction of Dr. Rachel Stutsman Ball.

TABLE 4

SOCIABILITY WITH OTHER CHILDREN

Scale value	Form *B*
25	Finds it difficult to approach other children and make friends.
105	Makes friends with any child who happens to be around him
56	Has a particular friend (Underline, own sex, opposite sex) whom he admires very much (hero-worship or crush)
87	Contributes to the ideas of the group though not a leader (cooperative companion)
30	Refuses to cooperate with other children unless he is the leader.
45	Hesitant in making suggestions to other children.
83	Interest of other children spurs him on to activity.
99	Has a pleasant manner of securing cooperation from other children.
78	Quarrels with other children only over serious issues
33	Picks on one particular child.
10	Rough and mean with other children.
17	Teases or torments younger children
29	Impatient with other children.
26	Enjoys seeing other children reprimanded.
23	Very critical of other children
93	Is sympathetic toward other children.
90	Tries to help the smaller children.
95	Forgiving of other children who have hurt him, taken his belongings, etc.
17	Seldom talks to other children
35	Cries easily in playing with other children.
99	Generous in letting other children share activities and possessions.
9	Selfish with other children; doesn't want to share possessions or let them enter into his activities.
52	Attention from other children leads him to "show off" or act silly
28	Doesn't want other children to get attention from adults.

As one method of studying agreement of ratings, average scores on sociability ratings made at different times were correlated. For 31 cases, the rank-difference correlation between ratings obtained four weeks apart was 67±.07. For an interval of eight weeks, the correlation was 55± 091, and for an interval of twelve weeks, .56±.084 For 27 children in nursery school over a two-year period, the rank-difference correlation between mean scores for the fall of 1930 and the spring of 1932 was .55± 096.

By pairing each statement with the one most like it in popularity and as near as possible in scale value, two equivalent forms of the sociability scale were made Several statements were eliminated as of little value. The equivalent forms, *A* and *B*, are given in Tables 3 and 4.

TABLE 5

Rater	No. children	r_{AB}
1	70	.80
2	69	.70
3	53	.59
4	53	.64

Table 5 shows the correlation between scores on Form *A* and Form *B* for each of four teachers. Teachers 1 and 2 were in the group for children from 3½ to 5 years of age, and teachers 3 and 4 in the group for children 2 to 3½. The ratings were made at three periods two weeks apart.

These figures show that the teachers who rated the younger children were less reliable than those who rated the older ones. They may have been poorer raters, but on the other hand it may be much more difficult to judge the social behavior of young children than that of children beyond the age of 3½ years

The average reliability for the four teachers is .68. If this is taken as the reliability of the unit test in the Spearman-Brown prophecy formula, the expected reliability of the whole scale (or for two raters) is .81; for three raters, .87; and for four raters, .90. If the average reliability for the two teachers of the older children is taken as that of the unit test, the reliability of the whole, or for two raters, is .86. Similarly, that for the younger children becomes .76.

The correlation of scores on Form *A* with scores on the whole scale is .93; of Form *B* and the whole scale, .90. These correlations are high enough to warrant using a shortened form, preferably Form *A*.

Relation to Age.

The product-moment correlation between 590 average scores and age (ranging from 2 to 17 years) is $-.21 \pm .027$, a negligible relationship. At all ages there is a piling up of scores in the upper half of the scale, with the median for successive age groups ranging from 65 to 79. Medians are shown in Table 6. Since the median Sociability score advances with age, one might expect the correlation between age and sociability to be slightly positive rather than negative ($r = -.21$), i.e., a slight tendency for the older children to be more sociable. It is within the range of possibility that the overlapping of

TABLE 6

Age in months	Median score
24- 35	65
36- 47	70
48- 59	73
60- 83	74
84-107	73
108-131	75
132-203	79

the age distributions might have enough inversions to give an *r* of —21, but one would not expect it. However, inversions in the lower percentiles (Table 35) indicate that this may be true.

While the coefficient of correlation between score and age is low, the percentile curves by age groups show the highest relationship with age of any of the nine scales This is the only scale for which percentiles are reported for more than four age groups. In the other eight scales the age differences were either not present or were so slight that age groups have been combined, and in some cases only one set of percentiles is given for the entire age distribution.

An analysis of the statements was made to find out if certain social behavior traits included in this scale were characteristic of different age levels The results show that the scale does not lend itself to age divisions.

ASCENDANCE-SUBMISSION

The Ascendance-Submission scale,[8] shown below (Table 7), consists of 48 items, many of which occur in pairs. For example, *"Opposition spurs him on to greater activity"* and *"Does not push the issue in case of opposition"* Some statements express varying degrees of a characteristic, such as *"Usually takes the initiative," "Can take the initiative if it is absolutely necessary," "Hesitates to initiate activity,"* and *"Helpless unless someone organizes activity for him."* Scale values for these four statements vary from 11 to 87. A fifth statement which has the very high scale value of 105 might be included in this group: *"Directs all activity about him."*

---

[8] The personality rating scales may be obtained from the Merrill-Palmer School, Detroit, Michigan. The set of nine scales including instruction sheet and scoring key is priced at ten cents with a 10 per cent reduction for twenty-five or more.

TABLE 7
Merrill-Palmer School Personality Rating

Date——————————

Name——————————Rated by ——————————

Directions for Rating:

Check only those statements which you feel are really true of the child. Do not guess if you are not reasonably sure. A few true statements are better than many half true ones.

ASCENDANCE-SUBMISSION

Scale value		
10	( )	Submits to any child who takes the initiative
7	( )	Even submits to *younger* children
31	( )	Submits to children of *his own age* (either sex)
68	( )	Dominates children more *immature* than himself.
84	( )	Dominates children of his own age (either sex)
101	( )	Dominates children *more mature* than himself
43	( )	Will submit to a *specific child* only.
51	( )	Submits occasionally to some other child
61	( )	Dominates a specific child only
71	( )	Submits to a leader only after a struggle to dominate.
49	( )	Is a follower in one *specific group* only
72	( )	Occasionally dominates a group
83	( )	Dominates a specific group only.
76	( )	Usually leads a small group
96	( )	Usually dominates a large group
96	( )	Decides who shall participate in the group activities
93	( )	Can organize the activities of a group to carry out a definite purpose.
112	( )	Is a leader in any group.
109	( )	Directs all activity about him.
56	( )	Leads or follows as the occasion demands.
60	( )	Neither leads nor follows, plays alone.
73	( )	Dominates other children by having greater material possessions which they covet
86	( )	Other children make many appeals to him for information
91	( )	Dominates other children through his ability to talk effectively.
96	( )	Other children appeal to him to make decisions for the group
98	( )	Dominates other children through their love or admiration for him
104	( )	Dominates other children through his wealth of ideas
83	( )	Definitely schemes to get others to carry out his plans
90	( )	Gives commands with an air of finality
11	( )	Helpless unless someone organizes activity for him
38	( )	Hesitates to initiate activity

TABLE 7 (*Continued*)

Scale value		
26	( )	Hesitates to make suggestions to other children.
33	( )	Usually follows the ideas of others for activity.
76	( )	Usually has his own ideas for activity.
63	( )	Can take the initiative if it is absolutely necessary.
88	( )	Usually takes the initiative.
34	( )	Seeks the approval of the leader before he acts.
40	( )	Does not push the issue in case of opposition.
42	( )	Stands aside to let others participate.
93	( )	Fights for his place as leader.
85	( )	Opposition spurs him on to greater activity
97	( )	Refuses to cooperate unless he is the leader
94	( )	Insists that other children do as he wishes.
19	( )	Does not defend his own rights with other children.
27	( )	Easily led into mischief by others.
81	( )	Much rivalry with other children.
52	( )	Fails to secure cooperation when he tries to direct activities
96	( )	Gets willing cooperation easily.

Reliability.

The Pearson product-moment coefficient of correlation between the ratings of two nursery-school teachers in November, 1934, is .73±.054 (30 children), and in April, 1935, .60±.087 (25 children). Teacher 1's ratings for November correlate with her April ratings .45±.12 (22 children), and Teacher 2's ratings correlate .51±.10 (25 children). The November average for the two raters correlates with the April average .45±.12 (22 children). Strictly speaking, these November-April correlations are not measures of reliability, for the scores at these two intervals may represent actual personality changes in the child over that period Thus they paint a darker picture of reliability than is justified.

While none of these correlations is high enough to justify the conclusion that this scale is a thoroughly reliable method of judging ascendance-submission, it compares favorably with the other scales. It presents a detailed picture of responses presumably ascendant or submissive in character, and brings into question the reliability of any judgments of ascendance-submission, particularly of young children. Lack of agreement on any of these scales may indicate that these children present different aspects of behavior to different judges, different situations, and at different times. This is a more reasonable interpretation than that the raters are less capable of checking these statements reliably because of defective ability to judge ascendance-submission.

INDEPENDENCE OF ADULT AFFECTION AND ATTENTION

A glance at this scale will show that the type of independence measured is emotional and not physical. These items give no indication of a child's ability to feed himself, wash his face, dress himself, climb to the top of the jungle gym, and so on. Rather, they attempt to evaluate his emotional independence of or dependence upon adults.

A copy of this scale, which consists of 33 items, follows (Table 8). As in the other scales, most of the items have opposites.

Miss Clare Atkinson (1), a Merrill-Palmer student, made a special study of this rating scale. Four sets of ratings made in the spring of 1930 by four different raters were analyzed for popularity of the statements, that is, the total number of times they were checked, and for their differentiating value.

The following tabulation shows the ten most popular and the ten least popular statements.

Ten Most Popular Statements

Pays no attention to visitors.
Independent of adult in having ideas about or planning work or play activities.
Leaves parent in matter-of-fact manner.
Independent of adult in overcoming difficulties
Matter-of-fact in relations with adults.
Perfectly natural in presence of adult
Not dependent upon adult for praise.
Bids for attention from adults.
Dependent upon adult to solve difficulties
Shows affection toward adults.

Ten Least Popular Statements

Makes friendly advances toward visitors
Presence of parent upsets child's regime
Seeks adult aid at every move.
Forms adult attachments often.
Cries to secure adult attention.
Asks for physical demonstration of affection
Emotional outburst upon leaving parent

TABLE 8

MERRILL-PALMER SCHOOL PERSONALITY RATING

Date————————

Name———————— Rated by ————————

Check only those statements which you feel are really true of the child. Do not guess if you are not reasonably sure. A few true statements are better than many half true ones

INDEPENDENCE OF ADULT AFFECTION AND ATTENTION

Scale value		
60	( )	Perfectly natural in the presence of adults.
22	( )	Always conscious of adult's presence
80	( )	Matter-of-fact in his relations with adults.
98	( )	Defiant attitude toward adults.
96	( )	Avoids adults as much as possible.
21	( )	Insists that a specific adult assist him
38	( )	Has great admiration for particular adults (hero worship).
101	( )	Independent of adult in overcoming difficulties.
11	( )	Dependent upon adult to solve difficulties
100	( )	Independent of adult in having ideas about or planning work or play activities
11	( )	Dependent upon adult for ideas and plans for work or play.
3	( )	Seeks adult aid at every move.
106	( )	Resents aid from adults.
86	( )	Pays no attention to visitors.
41	( )	Makes friendly advances toward visitors.
32	( )	Acts silly in presence of visitors or newcomers
30	( )	Presence of parent (underline: father, mother) upsets the child's regular routine.
75	( )	Leaves parent (underline father, mother) in a matter-of-fact manner.
14	( )	Emotional outburst upon leaving parent (underline: father, mother).
24	( )	Bids for attention from adults
15	( )	Cries often to secure adult attention.
19	( )	Does his best only when praised by adults.
92	( )	Not dependent upon praise from adult to do his best
25	( )	Seems worried that adults won't like him.
30	( )	Expects adults to feel sorry when he is not good
27	( )	Forms adult attachments often
10	( )	Craves and definitely seeks affection from adults
30	( )	Craves affection from adults but afraid to show it.
16	( )	Asks for physical demonstration of affection from adults.
22	( )	Gives physical demonstration of affection.
36	( )	Shows affection toward adults
46	( )	Usually affectionate but bursts occasionally into "I hate you"
108	( )	Resents affection from adults.

Seems worried that adults won't like him
Usually affectionate but bursts occasionally into "I hate you"
Acts silly in presence of visitors

Thus it can be seen that the typical nursery-school child tends to be fairly independent of adults, and since this is a much desired characteristic, it is not surprising to find the teachers discovering it frequently in the children.

One of the raters has been in the Merrill-Palmer nursery school during the entire seven-year period in which the scales have been in use. A comparison of her ratings with the results just described shows the first eight most popular statements the same in both lists, though in a slightly different order, and nine of the ten least popular statements the same in both lists.

By using median scores (1930 ratings), ratings of the 10 most independent children and the 10 most dependent children were selected for special study. The number of times each statement was checked was recorded for these two groups. A comparison of these frequencies with the total popularity and the difference between the two frequencies was then taken as a measure of the differentiating power of the individual statements. The following four non-differentiating statements were thus eliminated: Unconscious of adult's presence when engaged in some activity, attempts to convince adults of his originality, special accomplishments, etc., craves affection from adults but is afraid to show it; and leaves parent in a matter-of-fact manner. The last two of these statements have not been omitted from the scale used in this study but will be omitted in the next revision.

Reliability.

The remaining statements were then paired for scale value and popularity, and two forms of the scale were obtained in order to study reliability by the split half method. The scale very definitely breaks in the middle, there being no statements with scale values in the 50's, 60's, and 70's. This break may be due to the fact that this category is very likely not a primary trait or that the scale included too many paired opposites, thus offering only the two extremes of independence and no midway position. The two equivalent forms *A* and *B*, are given in Tables 9 and 10. The items are given in the order of ascending scale value to show when the scales break in the middle. Mean scores rather than medians must be used.

TABLE 9
INDEPENDENCE OF ADULT AFFECTION AND ATTENTION

Scale value	Form *A*
3	Seeks adult aid at every move.
10	Craves and definitely seeks affection from adults
11	Dependent upon adult to solve difficulties
12	Doesn't want other children to get attention from adults, wants it all himself.
15	Cries often to secure adult attention.
16	Asks for physical demonstration of affection from adults.
18	Very sensitive to attitude of adult toward him.
22	Always conscious of adult's presence
24	Bids for attention from adults.
27	Forms adult attachments often.
28	Attempts to talk to adults a great deal
30	Presence of parent upsets child's regular routine.
36	Shows affection toward adults
46	Usually affectionate but bursts occasionally into "I hate you."
*	
80	Matter-of-fact in relations with adults.
82	Child plays in usual manner when parents are present
94	Attitude of adult toward child seems to make little difference.
96	Avoids adults as much as possible.
101	Independent of adult in overcoming difficulties
106	Resents aid from adults
108	Resents affection from adults.

*Scale breaks in middle.

TABLE 10
INDEPENDENCE OF ADULT AFFECTION AND ATTENTION

Scale value	Form *B*
3	Seeks adult aid at every move
10	Craves and definitely seeks affection from adults.
11	Dependent upon adult for ideas and plans for work or play.
11	Tries to get adult to interfere and fight his battles for him rather than stand up for his own rights.
14	Emotional outburst upon leaving parents.
18	Fondles adults a great deal
19	Does his best only when praised by adults
21	Insists that a specific adult assist him.
24	Seeks help of adult as means of securing attention.
25	Seems worried that adults won't like him.
30	Expects adults to feel sorry when he is not good.
30	Seeks adult attention by appeals for information.
32	Acts silly in presence of visitors or newcomers.
38	Has great admiration for particular adults.

TABLE 10 (*continued*)

Scale value	Form *B*
41 *	Makes friendly advances toward visitors
82	Almost never initiates conversation with adults
86	Pays no attention to visitors.
92	Not dependent upon praise from adult to do his best.
98	Defiant attitude toward adults
100	Independent of adult in having ideas about or planning work or play activities.
106	Resents adult aid
108	Resents affection from adults

*Scale breaks in middle

The following tabulation (Table 11) shows the correlations obtained between Forms *A* and *B*

TABLE 11

Rater	No. children	r_{AB}
1	94	843
2	95	751
3	69	848
4	68	.645

The average of these four r's is 772. If this correlation is used as the unit test in the Spearman-Brown prophecy formula, the reliability of the whole scale would be .871. This may also be interpreted as the expected reliability for two raters using the short form. Our observed correlation for two rates, r_{AB}, is .847, which agrees fairly well with the expected reliability of .871.

Form *A* correlates .942 with the whole scale; Form *B*, 911. Form *A* is thus the better one to use for a shortened form Analysis of the two forms shows that there is a tendency for the less popular statements to be in Form *B*, and the statements of low scale value tend not to be popular in this form. Thus the mean is likely to be too high in Form *B*.

The correlations between scores by the two raters on the same children are .75 for raters 1 and 2 and .542 for raters 3 and 4. This rather low correlation is not very important, however, for part of the difference is due to an actual difference in the way the children respond to different teachers At least three teachers are quite consistent in their own ratings, as judged by r_{tt} The low r_{AB} for rater 4 is due to the fact that she marked very few statements for each child, which also accounts for the low correlation between rater 3 and rater 4

Relation to Age.

Ratings made in 1931 on a group of Merrill-Palmer children, a settlement nursery school, and a public school nursery were analyzed for age differences Over a period of three months 121 ratings were made on 82 children ranging in age from two to six years. The children were divided into three age groups, and the items of the scale were studied to find out if certain types of behavior tended to occur at different ages. The three pictures were strikingly similar and there were no differences great enough to be significantly differentiating. The product-moment correlation between chronological age and independence scores for the total group was —.004 Percentile curves were drawn by age groups, but since there were no regular variations, all age groups were combined.

Recommendations for Changes

Most of the statements of the scale can be checked on the basis of actual observations of behavior, things said or done, which indicate independence of or dependence upon adult affection and attention. However, there are a few statements which seem to depend more upon the subjective judgment of the observer than upon observations of specific behavior. For example: *"Perfectly natural in the presence of adults," "Defiant attitude toward adults," "Seems worried that adults won't like him,"* and *"Expects adult to feel sorry when he is not good."* It is suggested that the statement *"Perfectly natural in the presence of adults"* be omitted since analysis has shown it to be too general a statement, requiring little thought on the part of the rater and subject to wide variation in interpretation. The statement *"Expects adult to feel sorry when he is not good"* should be omitted since it is rarely used and is too general in interpretation. The statement *"Defiant attitude toward adults"* should be changed to *"Expresses defiance of adults"* thus suggesting that the observer base his judgment on specific instances of behavior rather than on general impressions For the same reason the statement *"Seems worried that* *adults won't like him"* should be changed to *"Frequently asks for adult's opinion of him"*

Form *A*, the shortened form of the scale, includes none of these statements However, we have not used this form for our ratings and consequently have no percentiles to publish for it

Tendency to Face Reality

Some persons seem to be able, at a very early age, to accept situations without evasion, while others cannot face an issue squarely even in adulthood. In abnormal psychology there are special terms to describe persons who constantly evade reality, compensate, and daydream This scale attempts to measure in children the tendency to face reality.

In the original mimeographed form of the scale there were 89 statements, a number much greater than seemed desirable Miss Frances C. Davis (5), a Merrill-Palmer student, made a study of this scale and reduced the number of items to 43 In order to find out what statements should be eliminated it was necessary to evaluate the separate items by a criterion of internal consistency.

If a child had been rated four or more times his records were used There were 302 ratings on 47 children These 47 children were arranged in rank order from highest to lowest on the basis of the median scores for all their ratings. The 10 children having the highest median scores were assumed to represent the extreme tendency to face reality, and the 10 having the lowest scores, the extreme tendency to evade reality The general uniformity of these two groups is indicated by the fact that the median percentile for the high group was above 82 and for the low group below 20

Differences between the high and low groups were computed for each item It was found that over half the statements were as likely to be checked for the high group as for the low one and thus could be considered non-differentiating. These 46 items were eliminated Scores on the shortened form were correlated with the long form with a coefficient of correlation of 97± 005 (47 cases).

The scale as used in the study is the abbreviated form. The high scale values indicate facing reality, the low values evading reality. A copy of the scale follows (Table 12).

While a student at the Merrill-Palmer School, Miss Winifred Reynolds (11) began a study which she later completed as a Master's

TABLE 12
MERRILL-PALMER SCHOOL
PERSONALITY RATING

Date——————

Name——————

Rated by——————

Directions for Rating:

Check only those statements which you feel are really true of the child. Do not guess if you are not reasonably sure A few true statements are better than many half true ones

Scale value		TENDENCY TO FACE REALITY
104	( )	Faces the issue squarely
90	( )	Concentrates his energy to accomplish a difficult task
91	( )	Meets situations in a quiet matter-of-fact manner.
88	( )	Has a wide range of constructive interests
84	( )	Does the hardest part of a task first
45	( )	Does the easiest part of a task first
29	( )	Dawdles to avoid a difficult task.
35	( )	Worries over things that may never happen.
78	( )	Does not worry over things that may never happen
42	( )	Worries over trivial matters.
68	( )	Does not worry over trivial matters.
35	( )	Feels unduly disappointed when his plans do not go right.
38	( )	Grieves too much over that which he has lost
91	( )	Accepts necessary facts as a matter of course
24	( )	Refuses to accept certain necessary facts.
94	( )	Does the best he can with what he has.
34	( )	Does not realize his own limitations
36	( )	Finds it difficult to accept his own limitations
44	( )	Not very self critical.
33	( )	Thinks he should always have the center of the stage.
82	( )	Recognizes and accepts the superiority of another child
89	( )	Accepts the fact that he can't excel in every type of activity.
88	( )	Appreciates and adjusts to a good authority.
28	( )	Does not appreciate or adjust to a good authority
36	( )	Attitude of defiance when not accepted by the group.
90	( )	Accepts just criticism willingly
29	( )	Finds it difficult to accept just blame for his faults.
25	( )	Expects to be excused from a difficult task.
21	( )	Regresses to babyish behavior in the face of difficulty.
27	( )	Expects to be pitied in the face of difficulty.
20	( )	Won't admit failure or defeat.
24	( )	Always has an alibi.
84	( )	Does not make excuses for his shortcomings.
28	( )	Becomes antagonistic when he fails.

Table 12 (*continued*)

Scale value		TENDENCY TO FACE REALITY
31	( )	Demands too much praise and bolstering to accomplish a great deal
64	( )	Expects to succeed at a task
87	( )	Quiet acceptance of success
88	( )	Knows when he has failed at a task
92	( )	Knows when he has done a task well
98	( )	Sees his faults or failures in their true light
42	( )	Tries to compensate for his shortcomings by excessive activity or animation
13	( )	Substitutes day-dreaming for real effort.
94	( )	Does not lose sense of reality even in imaginative play

thesis at Ohio State University under the title "The Rating of a Group of Preschool and Preadolescent Children on Their Tendency to Face Reality." She used the ratings of Merrill-Palmer nursery-school and recreational-club children and Ohio State University nursery-school children.

Reliability.

Table 13 shows the correlation between the scores of two raters

TABLE 13

CORRELATION

Group	No	Oct 15	Nov 5	Dec 3	Jan 21	Apr 14
I	18	.741±.08	.421±.14	.712±.07	.618±.05	.717±.08
II	25	.825±.04	.840±.04	.432±.12	.593±.09	.782±.05
O S U	26	.412±.12	.647±.08	.746±.06	.846±.04	.488±.13

Group I included the children from 2 to 3½ years of age, Group II children from 3½ to 5. The Ohio State University group included children from 2 to 5 years of age. The recreational clubs were made up of children from 5 to 14½.

The grouping for the older children shown in Table 14 indicates the age divisions as they met at the School.

TABLE 14

Group	No. children	Correlation Feb 1	Approximate age range, years
1	27	.140±.13	5- 6
2	22	.255±.14	7- 8
3	20	.748±.06	9-10
4	18	.308±.15	11-14

We have not made an analysis sufficiently fine to show the possible reasons for the variation from month to month in the ratings of the Merrill-Palmer children. The low correlation for the Ohio State University group on October 15 was probably due to the fact that one of the raters was entirely strange to the children at the beginning of the year. When the April 14 rating was made the group had been divided into two sections, which made each teacher's contact with part of the children less than it had been formerly. Some of the correlations for the nursery-school children are high enough to be considered fairly satisfactory for this type of data, but further study needs to be made of the reliability of the scales.

The low correlations for the older children are probably due in part to the fact that the leaders' contact with the children was weekly rather than daily and that each leader saw the children in specific activities rater than in all activities

Relation to Age

The correlation between age and scores for these groups of children was 092± 05 which indicates that the tendency to face reality, as measured by this scale, is not associated with age. The percentile curves show the same thing. Separate curves were drawn for five different age groups, but they followed so nearly the same course that they have been combined, and only one set of percentiles is given for the entire age range.

Analysis of Statements.

An analysis of the frequency with which the statements were checked was made to see which statements were marked most frequently and whether they tended to be characteristic of different age levels The statement *"Appreciates and adjusts to a good authority"* was checked the greatest number of times by all the raters at every age level The opposite of this statement, *"Does not appreciate or adjust to a good authority,"* was checked the smallest number of times

All ages combined, the 10 most popular and 10 least popular statements were the following:

Ten Most Popular Statements

Appreciates and adjusts to a good authority
Accepts necessary facts as a matter of course

Meets situations in a quiet, matter-of-fact manner
Accepts just criticism willingly
Recognizes and accepts the superiority of another child
Knows when he has done a task well
Concentrates his energy to accomplish a difficult task
Does not worry over things that may never happen
Faces the issue squarely
Accepts the fact that he can't excel in every type of activity.

Ten Least Popular Statements

Becomes antagonistic when he fails
Tries to compensate for his shortcomings by excessive activity or animation.
Expects to be pitied in the face of difficulty
Worries over things that may never happen
Does not realize his own limitations
Substitutes day-dreaming for real effort.
Finds it difficult to accept his own limitations
Attitude of defiance when not accepted by the group
Grieves too much over that which he has lost
Does not appreciate or adjust to a good authority

The 10 statements most frequently checked are items descriptive of a well-adjusted child who faces reality, and the 10 least frequently checked statements show just the opposite picture The same trend is shown in the percentiles, for the median score for the various age groups is between 70 and 75 Theoretically the mid-score should be 60. Thus it appears that these children tend toward good adjustment so far as this trait is concerned The results of this study of popularity of statements are similar to that obtained for the scale, Independence of Adult Affection and Attention

RESPECT FOR PROPERTY RIGHTS

In any group of children the ownership of toys and materials is of considerable concern, and children as well as nursery-school teachers are constantly confronted with the question of the possession, use, and care of property The scale of 30 statements (Table 16) attempts to measure the child's attitude toward property and his respect for property rights.

The scale was developed especially for use with preschool children, but has also been used with children up to the age of 13 years

A former Merrill-Palmer student, Miss Ethel Bickham (2), made a study of the ratings made by nursery-school teachers in two school years, 1930-31 and 1931-32. The subjects of her study were 45 Merrill-Palmer children ranging in age from 2 years, 2 months to 4 years, 10 months. They were rated three times the first year and five times the second year

Agreement between Raters.

Rater 1 and Rater 2 rated the older group of children, and Rater 3 and Rater 4 the younger group. Table 15 shows correlations between two raters rating at the same time, and between different ratings of the same rater at intervals of one month and four months.

TABLE 15

Rater	Interval	No. subjects	Correlation
R_1 with R_2	Same day	24	.58±.09
R_1 with R_1	1 month	23	.68±.08
R_1 with R_1	4 months	22	.71±.08
R_3 with R_4	Same day	15	.70±.09
R_3 with R_3	1 month	13	.63±.10
R_3 with R_3	4 months	11	—.12±.20

The correlations between raters are probably as high as can be expected for this type of data. With one exception, the self-correlations are high enough to indicate fair reliability. The low correlation between the two ratings of R_3 at a four-month interval may be due to the fact that the children rated were very young and new to the nursery-school situation at the time the first ratings were made and doubtless learned a great deal in the four-month interval. Thus the lack of correlation may be attributable to actual change in the children rather than to low reliability in the scale.

The ratings for 24 children in the nursery school during 1930-31 and 1931-32 showed a correlation after a year interval of .83±.04. This does not mean that there has been no gain in score but that the ranking has remained relatively constant. Twenty-three of the twenty-four children showed improvement. There was an average gain in score of 10.02 points. The three children who gained the most had an average increase of 18.53, and the three who gained the

TABLE 16
MERRILL-PALMER SCHOOL
PERSONALITY RATING

Date——————
Name——————Rated by——————

Directions for Rating

Check only those statements which you feel are really true of the child Do not guess if you are not reasonably sure A few true statements are better than many half true ones

Scale value		RESPECT FOR PROPERTY RIGHTS
82	( )	Distinguishes between his own and property of others.
4	( )	Takes toys belonging to other children from their lockers.
23	( )	Takes school property or that of other children home.
84	( )	Realizes that school property belongs at school
87	( )	Does not take possessions of other children without permission.
76	( )	Understands meaning of waiting his turn.
26	( )	Wants toy or play materials as soon as has desire regardless of whether or not it is his turn
88	( )	Takes good care of school property while using it
18	( )	Rough and destructive with school property.
25	( )	Tries to take equipment away from other children
29	( )	Wants to keep a particular piece of equipment even if not using it himself
78	( )	Gives up equipment to other children as soon as finished with it
34	( )	Usually desires equipment being used by others rather than what he has
24	( )	Resorts to slyness to get equipment being used by others
18	( )	Hits or knocks down child in attempt to get equipment from him
38	( )	Monopolizes certain pieces of equipment.
0	( )	Absolutely no sense of property values
110	( )	Extreme sense of property rights and keen desire to see them enforced.
5	( )	Extremely destructive of toys and equipment in school
105	( )	Shows extreme consideration for school property
106	( )	Shows extreme consideration for possessions of others
85	( )	Genuinely sorry when he has destroyed another's possessions.
5	( )	Takes pleasure in destroying possessions of others
46	( )	Impatient in waiting turn for a new toy
10	( )	Often takes school property or that of other children home to keep
76	( )	Takes good care of his own possessions
44	( )	Careless with his own possessions
88	( )	Takes good care of possessions of other children.
33	( )	Careless when using possessions of other children
12	( )	Plans to take possessions of others for himself even when he is told not to

least, 1 65 A study of these cases shows that the children with the lowest scores made the greatest gain; children with high scores gained very little. It is possible that the scale is too restricted at the upper end to allow for improvement. A child whose initial score is high may show improvement in his behavior but the scale may not be sensitive enough at the upper levels to measure this improvement.

Relation to Age.

The correlation between age and scores on this scale is .17± 10, which is too low to be of any significance

The percentile curves for age groups show a slight tendency to increase with age, but the tendency is not clear-cut at all levels However, the oldest group, from 9 to 13 years, had no score lower than 50, whereas all the other age groups began with a score of 25. The percentile curves represent a much wider age range than do the ratings used in the age correlation.

COMPLIANCE WITH ROUTINE

This scale consists of 31 statements about the various routines experienced by children in the nursery school (Table 17). It is so specifically in terms of nursery-school routine that it is applicable only to children of preschool age. Not as much work has been done on this scale as on some of the others, though the ratings have been continued over a period of six years.

Relation to Age.

No correlations with age have been computed, but the percentiles for the two- and three-year-old children are not sufficiently different to report them separately Beyond the twentieth percentile, the percentiles for four-year-old children were consistently higher than for the younger children, and have been reported separately. These percentiles suggest a slight relation to age.

Reliability.

In the early years of the study the reliability of this scale was not tested. However, correlations for the 1935 ratings are much like those obtained for the other scales, and it is probable that this scale is about as reliable as the others. The correlation between R_1 and R_2 for January, 1935, is 53±.11, and for the same raters in April of that year, 63± 072. The reason for the lower correlation in January

TABLE 17

Merrill-Palmer School
Personality Rating

Date——————————

Name——————————————Rated by——————————

Directions for Rating

Check only those statements which you feel are really true of the child Do not guess if you are not reasonably sure A few true statements are better than many half true ones

Scale value		COMPLIANCE WITH ROUTINE
99	( )	Adjusts immediately to the daily routine
11	( )	Objects violently to routine activities
94	( )	Always goes through the daily procedure willingly
83	( )	Usually goes through the daily procedure willingly
17	( )	Has to be constantly urged to carry out routine activities.
28	( )	Takes a long times to adjust to the daily routine
12	( )	Tries to prevent other children from carrying out the routine activities
95	( )	Quiet enjoyment from routine activities
88	( )	Accepts the routine as a matter-of-course
83	( )	Responds readily to direction in the day's routine
86	( )	Likes to assist the adult in routine tasks
50	( )	Cooperates or not in routine, according to his mood
38	( )	Acts silly at the lunch table
24	( )	Refuses many foods
32	( )	Often cries during nap period
27	( )	Objects to being examined by the nurse
26	( )	Talks and laughs with adjacent children during rest period or nap.
89	( )	Routine proceeds as usual in the presence of visitors
53	( )	Presence of visitors upsets his routine
16	( )	Presence of a specific child upsets his routine
26	( )	Emotional outburst upon leaving parents
97	( )	Business like and systematic in endeavoring to carry out routine activities
34	( )	Dawdles over routine activities
39	( )	Carries out routine tasks in a haphazard manner
93	( )	Makes a routine of his play activities
92	( )	Always cooperates in trying to keep the schoolrooms neat and clean
80	( )	Usually cooperates in trying to keep the schoolrooms neat and clean
77	( )	Usually puts things away carefully
16	( )	Occasionally puts things away
28	( )	Seldom cooperates in trying to keep the schoolrooms neat and clean
13	( )	Never puts things away

may be that these ratings were made the first week after the Christmas holidays, and the children may have been more difficult to rate at this time than they were later in the year.

Effect of Sequence of Statements.

Unlike the other eight scales, in which the statements are arranged in random order, the scale *Compliance with Routine* was constructed in two forms, with the statements identical in both, but arranged in Form *A* according to scale value, from lowest to highest; and in Form *B*, according to kinds of routine, and, within each group, of varying degrees of compliance with the type of routine in question. The scale published here is Form *B*.

As a special problem in mental growth, Miss Gladys Eesley (6) made a study of the effect of the sequence of statements on the child's score. Ratings made from September, 1931, to June, 1932, were used in the study Form *A* blanks were checked by two raters for each child in October, November, January, and May, and Form *B* blanks by the same raters two weeks later. At the time these ratings were made there were two separate nursery schools at the Merrill-Palmer School, one for children up to 3½ years and one for children 3½ to 5 years. Teachers *W* and *M* rated the younger children, and teachers *J* and *K* the older ones. Each teacher's ratings were grouped together for each of the two forms, regardless of the number of times a single child had been rated. Some children were rated four times on each form, and all these ratings were used. Ratings on Form *B* were used only if the teacher had rated the same child on Form *A* two weeks previously Table 18 shows the number of children rated by each teacher.

TABLE 18

Teacher	No children rated
J	75
K	99
W	49
M	58
	281

Each teacher's ratings, as well as the total ratings on Form *A*, were correlated with the ratings on Form *B*. The correlations are given in Table 19.

TABLE 19

Teacher	Correlation
J	.74±.04
K	.79±.03
W	.81±.03
M	.73±.04
All	.76±.02

These correlations are high enough to indicate fair agreement between scores on the two forms.

In order to see how well two successive sets of the same form agree, the October 15 and November 12 ratings on Form *A* were correlated The correlation was .53±.04 The correlation between the Form *B* ratings of October 29 and November 26 was .63±.05. Only three teachers were included in the last correlation, for teacher *M* did not rate this set. These correlations between the same forms with a month interval are smaller than the correlations of one form with another

The following figures (Table 20) show the average scores for Forms *A* and *B* for each teacher and for all four combined

TABLE 20
AVERAGE SCORES

Teacher	Form A	Form B
J	73.30	73.45
K	72.00	76.25
W	72.00	74.15
M	72.00	75.45
All	72.35	74.95

From these averages it can be seen that in all cases the ratings on Form *B* are higher than those on Form *A* When all ratings are put together, the average on Form *B* is 2.60 points higher than that on Form *A* There are 93 chances in 100 that the true difference between the two is greater than zero. There is a tendency, then, to rate children higher on Form *B*, the form on which the statements are not arranged in the order of their score value

To study the effect of position on the statements checked, a distribution was made of the statements checked for each form of the rating scale Rating sheets were included if both teachers had rated the child. Results show that there were only three statements with a difference greater than ten between the percentages checked on the two forms Two of these three statements were toward the lower end of the scale where the score values are highest. These two statements were checked more often on Form *B* than on *A*. Such statements may account for the slightly higher scores on Form *B*.

A study of the differences between the two forms shows that there is a slight tendency to check statements having smaller score values more frequently on Form *A* than on Form *B*, whereas the seven statements having the largest score values are checked more often on Form *B* This tendency of course makes the average score slightly higher on Form *B* than on Form *A* Thus, it is evident that the raters tend to mark more items indicating good compliance with routine when the statements are arranged according to type of routine and degree of compliance, than when the statements are arranged from lowest to highest score

RESPONSE TO AUTHORITY

Anyone who has observed groups of children knows that they show two widely divergent ways of responding to authority: that is, some children accede to all requests, and others are defiant of all authority. As in other aspects of behavior, there are variations between these two extremes and the majority of children fall in the middle group. The ability to adjust satisfactorily to authority may make or mar success and happiness in life It is essential not only to be able to accept the dictates of a superior, but also to know when an authority is a poor one and be able to resist it.

Learning what authorities to obey is an extremely important task for a preschool child. He has been accustomed to authority at home, and when he enters nursery school it is probable that he must learn to accept the authority of several other persons.

To study this important aspect of behavior, the scale *Response to Authority* was constructed In its present form it consists of 28 statements (Table 21).

Instead of being scored on the basis of 1 to 11, as in the other scales, the statements were ranked from those showing the least response to authority to those showing the greatest Since there were 44 statements, the scores ranged from 1 to 44. Some of the statements were later eliminated as unsuitable, and the number was reduced to 28. The percentiles presented in this study are based on the original scores but the revised scale values, computed on the 1 to

TABLE 21
Merrill-Palmer School
Personality Rating

Date——————

Name——————Rated by——————

Directions for Rating

Check only those statements which you feel are really true of the child. Do not guess if you are not reasonably sure A few true statements are better than many half true ones.

Scale values			
Original	Revised		RESPONSE TO AUTHORITY
17	35	( )	Attempts to change conversation from suggested activity to other channels
39	97	( )	Adds cooperative additions to the suggestion
3	1	( )	Resistant even against suggestion.
6	12	( )	Planned evasion
35	95	( )	Proud of his cooperation
18	41	( )	Lags in following suggestion
34	91	( )	Responds without undue delay to authority
5	14	( )	Cries if has to submit to authority.
4	11	( )	Runs away if called
42	104	( )	Comes quickly if called
10	26	( )	Thinks of immediate arguments as to why he shouldn't do suggested activity.
25	48	( )	Contemplates legitimate suggestion long time before doing it.
9	24	( )	Resists when required to do something new
31	60	( )	Says "No" but does suggested activity
13	31	( )	Says "Yes" but doesn't do suggested activity
7	13	( )	Frowns, shrugs shoulders, pouts or stamps foot when suggestion is made
23	45	( )	Resistant only when in a particular mood
11	28	( )	Pretends not to hear.
28	55	( )	So absorbed in own thoughts that doesn't comprehend
2	1	( )	Defiant against authority
44	110	( )	Accepts any command without question
21	42	( )	Experimentation with new authority to see how far he can go
1	0	( )	Rebels physically temper-tantrum, hitting, kicking, etc
12	30	( )	Pretended absorption to evade suggestion
38	96	( )	Cooperative and responsible
20	42	( )	Child tries to get task done by the person who suggests it
14	31	( )	Follows suggestion only while teacher is in sight
16	33	( )	Resists if suggestion is not about the things he has planned himself.

II scoring method, are published along with the others. Anyone using the scale should score by the revised scale values and compute new percentiles.

This scale needs further study, for it lacks statements between the values of 60 and 90.

Reliability of the Scale.

Miss Thelma F McClure (10), a Merrill-Palmer student, studied the reliability of this scale. The tabulation (Table 22) shows the rank-difference correlation between ratings for two sets of teachers. R_1 and R_2 rated the younger children, R_3 and R_4 the older ones From 19 to 25 children were used in each correlation.

TABLE 22

Raters	CORRELATION October	December
R_1 and R_2	75± 07	.66±.09
R_3 and R_4	37± 136	76±.067

The teachers who rated the younger children showed more consistency at the two periods of rating than did those who rated the older children. It is possible that the older children made more discrimination between the head and assistant teacher at the beginning of the year than did the younger children, and thus the two teachers tended to rate them differently in their response to authority Since response to authority is so largely a matter of the interaction of child and adult, there is likely to be a considerable difference of opinion about children's behavior in this respect. A child may respond readily to one adult's authority and at the same time be rather defiant of another's This difference is particularly noticeable in children who tend to have favorite teachers.

Relation to Age.

Six percentile curves for age groups from two to sixteen were computed, but since there are only very slight differences they have been combined in one curve. A coefficient of correlation for age and score has not been computed, but the chances are that such a correlation would be insignificant

ATTRACTIVENESS OF PERSONALITY

This scale is the longest of the nine, consisting of 75 items, most of which are short descriptive words or phrases. The other seven scales

TABLE 23
MERRILL-PALMER SCHOOL
PERSONALITY RATING

Date ____________________

Name ______________________________ Rated by ______________________

Directions for Rating

Check only those statements which you feel are really true of the child. Do not guess if you are not reasonably sure A few true statements are better than many half true ones

Scale value		ATTRACTIVENESS OF PERSONALITY
95	( )	Unusually happy disposition
82	( )	Nearly always smiling
94	( )	Smile lights up his whole face.
81	( )	Has a contagious laugh.
18	( )	Almost always seems unhappy.
25	( )	Almost never laughs or smiles.
1	( )	Extremely disagreeable manner
95	( )	Has an unusually good sense of humor
72	( )	Has a fairly good sense of humor
71	( )	Has a way of making an appeal with his eyes
86	( )	Has a pleasing manner of speech.
95	( )	Thoughtful of others
88	( )	Sympathetic nature
15	( )	Inconsiderate of others
45	( )	Not affectionate.
6	( )	Extremely selfish.
42	( )	Moderately selfish.
83	( )	Polite.
7	( )	Rude
60	( )	Mischievous
89	( )	Brave when hurt
20	( )	Very babyish when hurt
100	( )	Truthful
68	( )	Very persistent
34	( )	Gives up readily
32	( )	Makes excuses.
72	( )	Seldom cries
98	( )	A good sport
14	( )	A poor sport
29	( )	Domineering
10	( )	Deceptive
57	( )	Impulsive
42	( )	Very variable
86	( )	Very stable.
62	( )	Rough and ready
81	( )	Forgiving nature

TABLE 23 (*continued*)

Scale value		ATTRACTIVENESS OF PERSONALITY
12	( )	Very quarrelsome
15	( )	Very stubborn
26	( )	Wanders around aimlessly
36	( )	Self-conscious
30	( )	Nervous in manner, over-talkative, over-anxious, etc.
6	( )	Very negativistic
19	( )	Is on the defensive all the time
68	( )	Happy-go-lucky
64	( )	Very methodical
66	( )	Confides in adults
97	( )	Intelligently cooperative.
25	( )	Tries to show off or acts silly a great deal.
24	( )	Repels friendly advances.
16	( )	Sulks when not given his own way.
77	( )	Makes pleasant conversation with adults
84	( )	Makes an effort to help adults.
76	( )	Genuinely sorry when he has displeased an adult.
36	( )	Becomes too attached to certain adults
23	( )	Egotistical.
29	( )	Repressed.
104	( )	Unaffected, spontaneous, natural
19	( )	Unpopular with other children.
82	( )	Imaginative.
31	( )	Lack of imagination.
92	( )	Adapts easily to a new situation
84	( )	Eager to try new things.
35	( )	Not much interested in new activities
78	( )	Seems to have a plan for every minute
88	( )	Brimming over with ideas for activity
73	( )	Fairly enthusiastic in work or play.
86	( )	Vivacious.
22	( )	Displays no enthusiasm in activities
83	( )	Plays or works vigorously
62	( )	Moderately energetic.
31	( )	Passive.
21	( )	Listless
31	( )	Haphazard methods of work or play.
35	( )	Extremely timid physically
38	( )	Lack of self-confidence

attempt to measure specific aspects of personality, whereas this scale and the ninth, Physical Attractiveness, have as their object the evaluation of personality in its affective aspect. A copy of the scale is shown in Table 23.

Relation to Age.

Percentile curves for six age groups from two to sixteen years so nearly coincided that they were combined into one curve.

PHYSICAL ATTRACTIVENESS

This scale consists of 68 short descriptive phrases, some of them describing the child's physical characteristics, others, impressions the rater may have of these characteristics A copy of the scale is given in Table 24

TABLE 24
MERRILL-PALMER SCHOOL
PERSONALITY RATING

Date——————

Name——————Rated by——————

Directions for Rating

Check only those statements which you feel are really true of the child Do not guess if you are not reasonably sure A few true statements are better than many half true ones

Scale value		PHYSICAL ATTRACTIVENESS
92	( )	Good body proportions.
27	( )	Body proportions poor
14	( )	Serious deformity
26	( )	Legs bowed
86	( )	Well-shaped head
27	( )	Peculiarly shaped head
57	( )	Very wide forehead.
38	( )	Very long neck
31	( )	Very short thick neck
104	( )	Features strikingly beautiful
96	( )	Beautiful features
56	( )	Ordinary features.
36	( )	Homely features.
7	( )	Repulsive features
47	( )	Square jaw
43	( )	Very prominent cheek bone
39	( )	Broad nose
42	( )	Very long face
42	( )	Very thin lips
31	( )	Very thick lips.
40	( )	Protruding chin
29	( )	Receding chin.
28	( )	Crooked teeth.
27	( )	Face badly scarred
96	( )	Unusually pleasant facial expression
8	( )	Unusually disagreeable facial expression
92	( )	Exceptionally beautiful eyes
86	( )	Expressive eyes.

TABLE 24 *(continued)*

Scale value		PHYSICAL ATTRACTIVENESS
73	( )	Very large eyes.
57	( )	Ordinary eyes
39	( )	Eyes seem too far apart.
36	( )	Very small eyes
34	( )	Eyes seem too close together
27	( )	Eyes slightly crossed
24	( )	Expressionless eyes.
93	( )	Exceptionally beautiful hair.
84	( )	Hair is neat and clean.
21	( )	Care of hair is neglected
16	( )	Hair extremely unattractive.
92	( )	Beautiful smooth skin
34	( )	Coarse skin.
33	( )	Skin not clear.
31	( )	Sallow complexion.
85	( )	Rosy cheeks
96	( )	Looks very healthy
16	( )	Very thin and emaciated
92	( )	Stands erect
27	( )	Seems to have poor posture.
92	( )	Walks with ease and grace.
36	( )	Has a peculiar walk
88	( )	Soft musical voice.
83	( )	Soft musical laugh
53	( )	Ordinary speaking voice.
27	( )	Harsh voice
26	( )	Loud, harsh laugh
20	( )	Whining voice.
84	( )	Very attractive clothes.
69	( )	Meticulously neat about clothes.
57	( )	Wears ordinary clothes.
32	( )	Clothes do not fit well
23	( )	Very untidy about clothes.
73	( )	Body and clothes usually clean.
66	( )	Hands, face, and neck usually clean.
29	( )	Hands and face usually dirty.
22	( )	Usually has a dirty neck
17	( )	Body seldom clean
8	( )	Nose is usually running
2	( )	Body always has an offensive odor.

Eesley's Study.

As part of the requirements for the Master's degree at Wayne University, Detroit, Miss Gladys Eesley (7) wrote a thesis entitled, "A Study of Teachers' Ratings of the Attractiveness of Personality and Physical Attractiveness of Nursery School Children." She used the teachers' ratings of the Merrill-Palmer children on these two scales and in addition had the teachers rank the children on physical and personality attractiveness. She wanted to know how well a teacher's ratings of the children agreed with her ranking of them, what the relation is between physical attractiveness and attractiveness of personality as measured by these scales, and what the relation is between age and intelligence and these ratings.

From October, 1931, to November, 1932, the children were rated five times on these two scales. The May, 1932, ratings were used as the best ones, since the teachers knew the children well by that time and also had had a school year's experience in using the scale. Besides rating the children in May, each teacher was asked to rank the children with respect to physical attractiveness and attractiveness of personality. Only the older children were used in this part of the study, because the teachers of the younger children felt that they could not rank the children in order from least to most attractive. Teacher *A* was the head teacher and Teacher *B* the assistant. The head teacher is a permanent member of the staff, the assistant usually stays only one year.

Each teacher's rank judgments of 24 children were compared with her scaled ratings of the same children. Rank-difference correlations were obtained as shown in Table 25.

TABLE 25

Teacher	Attractiveness of personality Correlation
A	.63± 08
B	66± 09
	Physical attractiveness Correlation
A	51± 10
B	09± 13

These results show higher correlations between ratings and judgments of attractiveness of personality than of physical attractiveness. For physical attractiveness, Teacher *A* showed fair agreement in the two methods, but Teacher *B* showed none at all. This result is surprising, for it would seem easier to judge physical attractiveness than that more intangible quality, attractiveness of personality.

Even though an attempt was made to make the two scales measure different qualities, there is evidently overlapping between the two. For this group of children, the correlation between the two scales was 46± 07 This figure agrees rather closely with that obtained for a single age group of children 48 to 59 months of age, where the correlation is 43± 057 (See Table 29).

Attractiveness of personality correlates with age only —14±.09, which result is in agreement with the results of the percentile curves computed for age groups. The relation between physical attractiveness and age is a little higher, i.e., —.48±.07. The six percentile curves are best combined into two groups, the youngest group having a higher median score than the older. These results indicate a tendency to judge younger children as more attractive than older ones.

The correlations with intelligence are much the same for the two scales, i.e., .32±.09 for attractiveness of personality, and 35± 08 for physical attractiveness Thus there is only a slight tendency to rate the more intelligent children as the more attractive.

PARENTS' RATINGS

While Mrs. Mary Edmunds Gerlaugh (8) was a student at the Merrill-Palmer School, she made a study to evaluate parents' ratings of pre-school children on six of the nine rating scales: Ascendance-Submission, Physical Attractiveness, Attractiveness of Personality, Independence of Adult Affection and Attention, Response to Authority, and Tendency to Face Reality. These six were chosen because they could be rated by parents in a home situation.

The ratings were made during the first semester of 1931 on children who attended the Merrill-Palmer nursery school. Ratings for 40 children were returned The number of parents rating on the different traits is shown in Table 26.

TABLE 26

Scale	No. fathers	No mothers
Ascendance-submission	31	36
Physical-attractiveness	35	40
Attractiveness of personality	35	40
Independence of adult affection and attention	35	40
Response to authority	31	39
Tendency to face reality	32	30

The father and mother of the child were requested to rate independently. The nursery-school teachers rated the children at the same time the parents rated them The children's ages ranged from 2 years to 4 years, 11 months

Average scores were computed for each scale for fathers, mothers, and nursery-school teachers, and comparisons were made between them.

Table 27 shows the percentage of parents rating their children lower than the teachers did

TABLE 27

Scale	Percentage of parents rating lower than teachers
Ascendance-submission	26 6
Attractiveness of personality	20 0
Independence of adult affection	88 5
Physical attractiveness	43 0
Response to authority	76 6
Tendency to face reality	74 2

Results show a rather consistent tendency for parents to rate their children higher than teachers did on ascendance-submission These parents doubtless considered leadership a desirable quality, and observation showed that they encouraged their children in its manifestation. The parents showed by their ratings that they thought their children expressed qualities of leadership at home, whereas the teachers had much more opportunity to observe each child in a larger group of children, in which the majority of children tend to express themselves more or less freely

On physical attractiveness there was fairly close agreement between parents and teachers There was less difference betwen parents' and teachers' ratings on this scale than on any other.

The difference between mean scores for parents and teachers on attractiveness of personality was a few points higher than on physical attractiveness, but the difference is not particularly significant. The trend was for parents to rate their children higher than the teachers did.

On response to authority 76.6 per cent of the parents rated their children lower than the teachers did This would indicate one of two things; that children respond more readily to authority at school than they do at home or that teachers judge the same behavior as more compliant than the parents do

The parents rated independence of adult affection and attention much lower than the teachers did, 88 5 per cent of the parents rating lower than the teachers.

The greatest difference between parents' and teachers' ratings occurred on the scale Tendency to Face Reality, the difference between fathers' and teachers' ratings being 18.49 points. Seventy-four per cent of the parents rated lower than the teachers. It is probable that the regular routine, consistent discipline, and companionship with contemporaries found in the nursery school help children to face facts as they are and learn to make adjustments much better than it is possible for them to do in many homes.

As judged by average differences between mean scores for six scales, the father and mother of the child were on the whole in closer agreement with each other than the two nursery-school teachers were. It is possible that the parents did not rate as independently as the teachers did, though they were instructed to rate the child without consulting each other.

However, rank-difference correlations between fathers' and mothers' scores show considerable variation in agreement for the various scales The correlations are given in Table 28.

TABLE 28

Scale	No. children	Correlation
Ascendance-submission	30	.78± 049
Attractiveness of personality	35	24± 107
Independence of adult affection	35	.51± 085
Physical attractiveness	35	.19±.110
Response to authority	30	.22±.117
Tendency to face reality	31	.51±.092

Three of the six correlations are so low in relation to the probable error that they are of no significance. Only one of the six scales, that on ascendance-submission, shows high agreement.

INTERCORRELATION OF THE NINE SCALES

To see what relation exists between the scales, each was correlated with every other one. In order to eliminate the factor of age, the ratings of children within one age group, 48 to 59 months, were studied Table 29 shows the intercorrelations (product-moment).

The most striking result of the intercorrelations is the low relation between the Ascendance-Submission scale and every other, all

SCALE	Sociability	Ascendance-submission	Tendency to face reality	Response to authority	Compliance with routine	Respect for property rights	Independence of adult affection	Attractiveness of personality	Physical attractiveness
				CORRELATION					
Sociability		10±.072	56± 049	45± 057	.40±.070	45± 068	37± 065	70± 036	35± 062
Ascendance-submission	10± 072	.	—07± 072	—03± 073	—.10± 079	.17±.073	20± 072	17± 071	.08± 072
Tendency to face reality	56± 049	—07± 072		62± 043	62± 049	45± 060	50± 054	49± 054	23± 067
Response to authority	.45± 057	—03± 073	62± 043		.75± 035	74± 034	31± 066	44± 057	13± 070
Compliance with routine	40± 070	—10± 079	.62± 049	75± 035		74± 037	.28± 074	28±.074	.25± 076
Respect for property rights	.45± 068	.17± 073	45± 060	74± 034	74± 037		37± 065	35± 066	55± 050
Independence of adult affection	37± 065	20± 072	50± 054	31± 066	28± 074	37± 065	.	—25± 068	17± 071
Attractiveness of personality	70± 036	17± 071	49± 054	44± 057	28± 074	35± 066	—25± 068		.43± 057
Physical attractiveness	35± 062	08± 072	23± 067	.13± 070	25± 076	55± 050	17± 071	43± 057	.

*(Based on from 66 to 95 children whose ages ranged from 48 to 59 months)

r's being 20 or below. It may be that here we are measuring a trait which is independent of the others.

The most closely related scales are *Response to Authority, Compliance with Routine,* and *Respect for Property Rights,* these correlations being in the 70's. It is possible that these three scales should be revised to form a single scale measuring a general trait of compliance. The only other scales which correlate to the extent of .70 are Sociability and Attractiveness of Personality. Apparently those children rated as sociable are likely to be rated as attractive.

The scales *Tendency to Face Reality, Response to Authority,* and *Compliance with Routine* show some relation with one another, the correlations being .62. It is possible that *Tendency to Face Reality* has some of the same elements as have the scales which seem to measure the general characteristic of compliance, and that learning to accept failure and success, praise and blame, and realizing one's limitations—in other words, facing reality—is very much akin to learning to respect authority and to be a responsible member of a group, so far as the daily routines go.

Much work remains to be done on these scales, for example, a factor analysis should be made to determine what fundamental traits are included in this series of rating scales. The scales would be considerably improved if some of the alterations suggested were made. New combinations of some of the scales might be tried, and possibly would prove more satisfactory tools for evaluating the fundamental traits than are the present scales.

GRAPHIC ANALYSIS

Even a cursory examination of the scores for individual children shows that there is a great deal of variability in the ratings. Raters are likely not only to disagree with each other when they rate a child at the same time but also to disagree in their own ratings at different times. It is also possible that having several persons rate a child over a period of years may make for greater variablity than having one person do all ratings. To study these variabilities, ratings on each of the nine scales were graphed for twenty-five children ranging in age from 1 year, 9 months to 5 years. The selection of cases was random, the first 22 children on the list being taken for study, with the addition of three other children who had been in nursery school from the ages of two to five years. During the time these children were in the nursery school there were two head teachers. Since one of the teachers has been in the school for all seven years of the study, only those ratings in which she was one of the pair of raters have been analyzed. The second rater in each case was the assistant teacher, who was usually a different person each year. Thus a comparison can be made between the ratings of a rater who has had a great deal of experience in working with the children and in using the scales, and those of a rater who is comparatively new both in the nursery school and in rating experience.

Table 30 gives the average number of ratings and assistant raters for each child by scales.

TABLE 30

Scale	Average No. ratings	Average No. assistant raters
Ascendance-submission	5.54	2.41
Attractiveness of personality	3.75	1.63
Compliance with routine	2.52	1.38
Independence of adult affection	3.95	2.58
Physical attractiveness	4.08	1.86
Respect for property rights	4.18	1.81
Response to authority	5.08	2.33
Sociability	5.54	2.62
Tendency to face reality	4.08	1.81

Means, standard deviations, and coefficients of variation were computed for the constant head teacher and for the varying assistant teachers for each of the nine scales. The results are given in Table 31. The means are based on actual scores rather than percentiles.

The coefficients of variation show that there is no constant tendency for the head teacher to be more variable than the assistant teachers, or *vice versa*. In three of the scales, *Respect for Property Rights, Response to Authority,* and *Tendency to Face Reality,* the head teacher was more variable than the assistant teachers. In four scales, *Ascendance-Submission, Attractiveness of Personality, Independence of Adult Affection,* and *Sociability,* she was less variable. In two scales, *Compliance with Routine* and *Physical Attractiveness,* one was as variable as the other. Thus it is evident that the variability of a single rater is very much like the variability of several raters.

The figures show also that the least amount of variation for both

TABLE 31

MEANS, STANDARD DEVIATIONS, AND COEFFICIENTS OF VARIATION FOR PERSONALITY RATINGS MADE BY HEAD TEACHER AND ASSISTANT TEACHERS ON 25 CHILDREN FOR WHOM GRAPHS WERE MADE

Scale	Head teacher Mean	Head teacher S.D.	Assistant teachers Mean	Assistant teachers S.D.	Coefficient of variation Head teacher	Coefficient of variation Ass't teachers	$*\frac{V_{ht}}{V_{at}}$
Ascendance-submission	65.39	10.89	58.57	16.22	16.6	27.71	.59
Attractiveness of personality	68.69	8.76	72.08	10.56	12.75	14.65	.87
Compliance with routine	74.50	13.40	74.95	13.62	17.98	18.17	.98
Independence of adult affection	67.35	13.97	65.38	14.90	20.74	22.78	.91
Physical attractiveness	75.05	7.995	75.57	7.995	10.65	10.57	1.007
Respect for property rights	74.20	17.50	76.84	13.37	23.58	17.38	1.36
Response to authority	25.67	8.04	28.50	7.10	31.32	24.91	1.25
Sociability	65.45	14.11	73.01	19.31	21.55	26.44	.81
Tendency to face reality	66.12	14.75	74.57	13.90	22.30	18.63	1.19

*Coefficient of variation of head teacher divided by coefficient of variation of assistant teachers the head teacher and the assistant teachers was in the scale *Physical Attractiveness.*

An inspection of these 225 graphs shows that even though there is considerable variation in the ratings, some of these curves tend to follow the same general pattern, sometimes on the same level and sometimes on different levels. Table 32 shows the percentage of curves which follow the same general pattern and those in which the agreement throughout the rating period was within 20 percentile points.

TABLE 32

Scale	Per cent Same general pattern	Agreement within 20 percentile points
Ascendance-submission	17	4
Attractiveness of personality	32	18
Compliance with routine	43	43
Independence of adult affection	33	21
Physical attractiveness	17	22
Respect for property rights	36	27
Response to authority	17	13
Sociability	29	13
Tendency to face reality	39	22

The percentage of graphs following the same general pattern on any level is less than 50 per cent in each of the nine scales. The results are based upon the subjective judgment of one person The other classification, agreement within 20 percentile points, is objective, for it is based on actual numbers. These percentages in seven of the nine scales are less than those for the same general pattern, indicating that for some of the cases the ratings for the two teachers were on different levels, even though they followed the same upward and downward trends.

This variability of ratings leads to several possible conclusions: (*a*) The scale items are too ambiguous to be used with the same meaning by different raters, (*b*) the method of checking only those items the rater thinks are really true of a child is poor, because some raters check many more statements than others; (*c*) the raters need more specific training before they are asked to rate children; or, (*d*) children are variable according to the situation and the person to whom they are responding, and thus personality measurements are variable This last point of view is in line with May's (9)

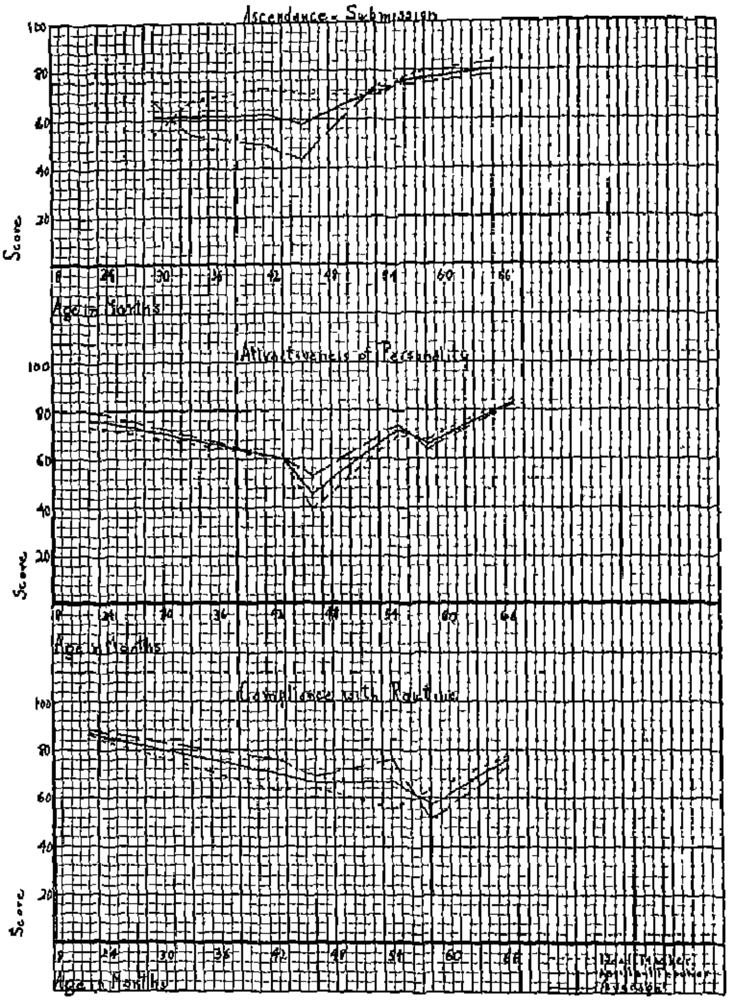


FIGURE 1a
GRAPHS COMPARING TEACHERS' RATINGS OF ONE CHILD ON THE NINE SCALES

definition of personality as a person's social-stimulus value. Each of these four possibilities needs further investigation and probably there is some truth in each of them. But certainly if these raters are at all typical of observers in general, all the results of this

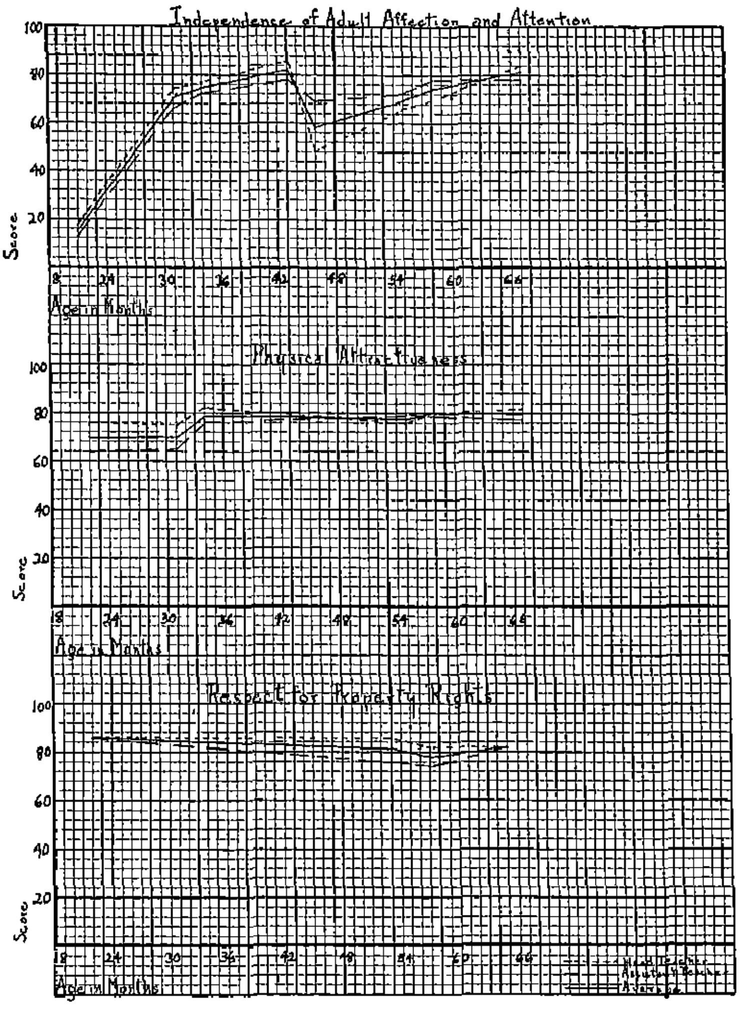


FIGURE 1b

Graphs Comparing Teachers' Ratings of One Child on the Nine Scales

study point to the fact that even well-trained observers who are with the children constantly and presumably know them quite well disagree not only in their ratings but also in their descriptions of the children's behavior. In other words, they interpret the child's

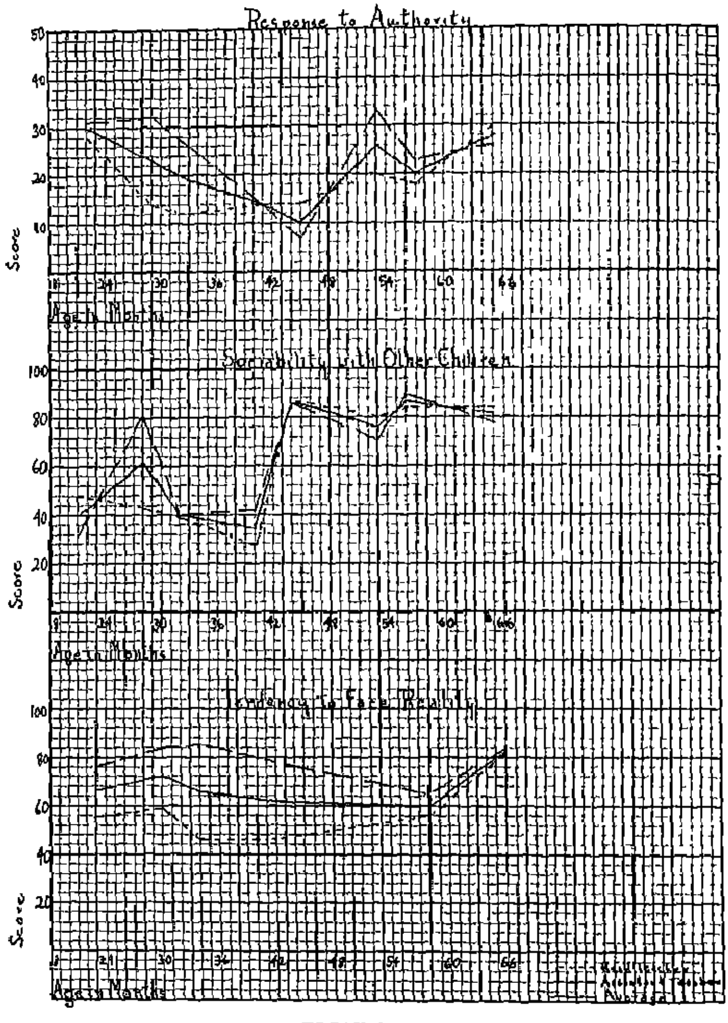


FIGURE 1c

GRAPHS COMPARING TEACHERS' RATINGS OF ONE CHILD ON THE NINE SCALES

behavior in the light of their experience with him, an interpretation which is colored by their personal relations with him. But it is characteristic of people that they respond in different ways to different persons, one person bringing out all the best qualities and another all the worst. Some teachers and parents seem to have the ability to smooth out all the difficulties children present, while others seem to create problems We are constantly interpreting other people's behavior in the light of our knowledge of them, and we are often surprised to find that our impression does not correspond at all with that of another person. Consequently, when two teachers rate a child independently of each other, perhaps we should not expect very close agreement when it is a matter of rating personality traits, for the subtle relations between people are extremely difficult to account for These graphs also show that some children are much more likely to be variable than others, a finding which is in accord with that of Conrad (4)

Some of the agreements and discrepancies in rating can be illustrated by a sample record The scores of this child on each of the scales are given in tabular and graphic form (Figure 1 and Table 33)

TABLE 33

Age	Rater	Score	Ave score	Percentile
		ASCENDANCE-SUBMISSION		
2-5	D_1	54	61	53
	H	68		
2-9	D_1	68	61	53
	H	54		
3-5	D_1	73	62	56
	D_2	50		
3-9	D_1	71	58	43
	D_2	44		
4-5	D_1	70	72	63
	D_2	74		
4-9	D_1	79	76	74
	D_2	74		
*5-5	M	78	80	84
1-10	D_3	83		
		ATTRACTIVENESS OF PERSONALITY		
1-10	D_1	74	77	No percentiles
	G	80		
3-7	D_1	60	60	19
	T	60		
3-10	D_1	40	46	4
	T	53		
4-7	D_1	70	72	49
	D_2	74		
4-10	D_1	68	66	32
	D_2	64		
5-7	M	83	84	
	D_3	85		90

*The last pair of ratings for each scale were the first made after the child entered the recreational club

TABLE 33 (*continued*)

Age	Rater	Score	Ave score	Percentile
		COMPLIANCE WITH ROUTINE		
1-10	D_1	86	87	No percentiles
	G	88		
3-6	D_1	63	70	34
	D_2	76		
3-10	D_1	64	66	26
	D_2	69		
4-6	D_1	56	67	21
	D_2	78		
4-10	D_1	63	58	14
	D_2	52		
5-6	M	73	76	No percentiles
	D_3	78		
		INDEPENDENCE OF ADULT AFFECTION AND ATTENTION		
1-9	D_1	18	16	No percentiles
	G	13		
2-7	D_1	74	70	58
	B	66		
2-10	D_1	77	74	70
	B	72		
3-7	D_1	86	82	90
	D_2	78		
3-10	D_1	48	58	28
	D_2	69		
4-7	D_1	66	69	55
	D_2	72		
4-10	D_1	69	74	70
	D_2	78		
5-7	M	78	81	88
	D_1	84		
		PHYSICAL ATTRACTIVENESS		
1-10	D_1	76	70	No percentiles
	G	64		
2-7	D_1	75	70	29
	W	65		
2-10	D_1	82	79	70
	W	76		
4-7	D_1	76	78	65
	D_2	79		
4-10	D_1	80	80	74
	D_2	79		
5-7	M	78	80	86
	D_3	82		

TABLE 33 (*continued*)

Rater	Score	Ave. score	Percentile
RESPECT FOR PROPERTY RIGHTS			
D_1	86	86	No percentiles
G	86		
D_1	86	81	60
D_2	76		
D_1	82	78	50
D_2	74		
M	82	82	60
D_3	82		
RESPONSE TO AUTHORITY			
D_1	28	30	No percentiles
G	31		
D_1	14	23	28
B	32		
D_1	12	19	18
B	26		
D_1	14	14	8
T	14		
D_1	14	10	3
T	7		
D_1	20	26	36
D_2	33		
D_1	18	20	20
D_2	23		
M	26	28	43
D_3	30		
SOCIABILITY WITH OTHER CHILDREN			
D_1	48	40	No percentiles
G	31		
D_1	43	62	46
B	81		
D_1	39	40	12
B	40		
D_1	27	35	3
T	42		
D_1	87	86	85
T	86		
D_1	80	76	58
G	71		
D_1	84	86	86
D_2	89		
M	78	81	70
D_3	84		

TABLE 33 (*continued*)

Age	Rater	Score	Ave score	Percentile
		TENDENCY TO FACE REALITY		
1-11	D_1	55	66	No percentiles
	G	76		
2-6	D_1	59	72	50
	B	84		
2-10	D_1	46	66	37
	B	85		
4-6	D_1	46	62	29
	D_2	78		
4-10	D_1	56	60	25
	D_2	65		
5-6	M	84	83	80
	D_3	82		

ANALYSIS OF RESULTS OF RATINGS ON INDIVIDUAL CHILD

For this child some of the scales show extremely close agreement, others quite divergent results. Those scales showing similar patterns are *Attractiveness of Personality, Independence of Adult Affection and Attention, Physical Attractiveness,* and *Respect for Property Rights* The other five show some disagreement and variability from time to time

For all children, as rated on all the scales, there is variability from one time of rating to the next; that is, changes in the average rating from one time to the next, not lack of agreement between raters Practically no child goes along on the same even level. A few show a consistent trend either upward or downward, but the majority show now an increase in the score, now a decrease.

Ascendance-Submission.

In an attempt to find out the reasons for such discrepancies in the *Ascendance-Submission* scale, the individual statements checked were studied. While the first two sets of ratings on ascendance-submission do not exactly agree, they are within twenty score points of each other. But the next two ratings, made when the child was 3 years, 5 months old and 3 years, 9 months old, show greater divergence. At 3 years, 5 months, raters D_1 and D_2 rated her. Rater D_1 made the notation that the child had been absent so much that the rating was not reliable. This fact alone could account for the difference.

But at 3-9 the raters agreed on but one statement, that the child usually had her own ideas for activity. Rater D_1, the head teacher, checked statements showing that the child dominated children more immature than herself but submitted occasionally to some other child, that she was a follower in a specific group, only occasionally dominating a group; that though she ordinarily neither led nor followed, preferring to play alone, she could lead or follow as the occasion demanded and organize the activities of a group of children to carry out a definite purpose; that when she did dominate, she did it by her wealth of ideas, easily getting willing cooperation

Rater D_2 thought that the child submitted to children of her own age; that she hesitated to initiate activity or to make suggestions to other children, that she could take the initiative if necessary, standing aside to let others participate.

Thus the two raters saw the child as showing different degrees of ascendance, D_1 picturing her as more ascendant than did D_2 D_1 has had more years of experience in rating and her ratings are likely to be more detailed than those of D_2, but her ratings are likely to be as inconsistent as those of any other rater. However, in this instance the difference in rating may be due to the fact that D_2 had been a student the previous year and was new in the teaching situation. The last two ratings agree very closely, and the statements checked indicate that the child is slightly above the median score in ascendance.

Attractiveness of Personality

While the agreement between raters on attractiveness of personality is fairly close, it is quite apparent that in the teachers' estimation the child became less attractive and then improved again When she was first rated at the age of 1 year, 10 months she had a smile that lighted up her whole face, was very persistent, seldom cried, was intelligently cooperative, unaffected, spontaneous, and natural, she adapted easily to a new situation, had an unusually happy disposition, a contagious laugh, and an unusually good sense of humor She was not affectionate, was moderately selfish, mischievous, brave when hurt, a good sport, very variable, and stubborn She was happy-go-lucky, made pleasant conversation with adults, and was eager to try new things

At 3 years, 10 months both raters scored the child as less attractive, D_1 rating her lower than T Both raters agreed that a smile lighted up her face, that she had a way of making an appeal with her eyes, was moderately selfish and deceptive, very variable, stubborn, and negativistic. D_1 thought that she almost always seemed unhappy, had a fairly good sense of humor, was not affectionate, was very babyish when hurt, a poor sport, very persistent, self-conscious, on the defensive all the time, and that she sulked when she was not given her own way. On the other hand, T checked the statements indicating that the child had a contagious laugh, was polite, mischievous, intelligently cooperative, and egotistical, made excuses, and conversed pleasantly with adults; that she was very stubborn and negativistic, and at the same time intelligently cooperative. When such contradictions occur it would be very helpful to have specific examples of the behavior by which the child was judged.

The ratings of the child at 4 years, 7 months and 4 years, 10 months show close agreement in score for the two raters. The descriptive statements written at the same time agree in portraying the child as attractive.

Age: 4-7
Rater· D_1

An attractive child because of her intense interest in such things as stories and handwork. She will stay with these activities for long periods of time. When she is pleased her face glows with happiness.

Her most disagreeable traits are diminishing: Resistance to adult directions and a little deceitfulness in things which she does not want to do

At times, she entertains other children by being a little silly, but this is a phase of development.

Rater: D_2

S. smiles and laughs a great deal and is one of the first to see the funny part of a story or situation. She enjoys teasing others. She is persistent in work and does not give up easily This shows up her stubbornness, for if she has some idea to carry out, she goes through with it, no matter what the adult says. She is usually busy and very imaginative.

Age· 4-10
Rater· D_1

> A much more charming child than she used to be Intelligently cooperative, sense of humor, has learned to tease and take teasing.
>
> Her disposition is a little variable Sometimes she is friendly and at other times rather sullen

Rater· D_2

> *S* has an attractive smile and very deep dimples She laughs a good deal, partly because of her silly behavior and partly because of her sense of humor She can be rude and at times is very inconsiderate of others It depends upon the mood she is in, for sometimes she is very considerate On the whole, she is a variable child She talks easily with adults She is a good sport and will take turns and settle other children's difficulties She has her own ideas about play and can amuse herself with very little material She loves to tell stories to herself and is very imaginative

Compliance with Routine

The scores on compliance with routine show fairly close agreement between the two raters until the age of 4 years, 6 months, both raters checking some statements that show the child as adjusting fairly easily to the nursery-school routine. But at 4 years, 6 months, D_1 rates her much lower than D_2, who said that she adjusted to the daily routine immediately, accepting it as a matter of course, while D_1 thought that she cooperated or not, according to her mood. For the first time the presence of a specific child upset her routine and she dawdled over routine activities and carried them out in a haphazard manner

The descriptive accounts written by the teachers to accompany the ratings are given to show that their opinions of the child are essentially the same.

Age· 4-6
Rater. D_1

> *S*'s attitude toward routine is somewhat improved. For a time it varied a great deal with her moods and with the person who was in charge of the routine, but recently she seems to be more stable; and because she has more skill in activities

> and doesn't dawdle as much, she gets through them with dispatch Generally enjoys assisting the adult.
>
> Socially more interested in her neighbors.

Rater: D_2

> *S* very seldom dawdles over routine, as she used to do, but seems to enjoy it and tries to be one of the first to get her outdoor clothes off or on, and is quick in the washroom. She is restless at nap and rest times, and is the only one of the older group who attempts to talk to other children at this time She is neat and orderly and enjoys putting paper, pencils, and scissors back in the cupboard
>
> She is very silly at times at the lunch table and laughs out loud and leads others to do the same

The ratings at 4 years, 10 months indicate that the child is not yet accepting routine tasks with ease. D_1 indicates that she has improved slightly, but D_2 indicates a decided drop The descriptive accounts indicate that the child has something yet to learn about complying with daily routines

Age: 4-10
Rater: D_1

> She is bored by most of the routine and has not learned to accept it as a thing that is just done in spite of one's interest or lack of it.

Rater: D_2

> *S* is very haphazard in all routine. She tries to escape from it when in certain moods and needs reminding about putting away toys. She has acted silly at lunch and is now at my table, where she is conforming She is using the entry way for her nap room, as she is very restless and disturbs the other children She tries to get other children to act the same way

It is interesting to note that during the three years this child was in the nursery school her compliance score was always below the median and her percentile score consistently decreased from the age of 1 year, 10 months to 4 years, 10 months. Since she has entered the recreational club her rating has increased but is not yet as high as her first rating was.

Independence of Adult Affection and Attention

The graphs of the scores for the two raters show fairly close agreement until the age of 3 years, 10 months These graphs show plainly that when the child was first rated at the age of 1 year, 9 months she was fairly dependent upon the teacher, but that in less than a year she had become fairly independent. Then at 3 years, 10 months she became more dependent again, D_2 rating her as more dependent than does D_1. D_2 thought she showed a defiant attitude toward adults but was independent of them in overcoming difficulties, even resenting aid She did, however, bid for attention from adults. D_1 said that she insisted upon having a specific adult to assist her, was dependent upon an adult to solve her difficulties, and did her best only when praised by adults At the time this dependence was shown, the child's baby sister was requiring considerable attention at home, and the child's seeking adult attention may have been a reflection of this home situation Since the age of 3 years, 10 months, the child has steadily become more independent of adults, her percentile score increasing from 28 to 88

Physical Attractiveness

Throughout the years of rating physical attractiveness there was fairly close agreement between the raters, and the level of the scores was high enough most of the time to indicate that the child was at least average in physical attractiveness One of her most recent ratings puts her in the upper quartile. The teachers' descriptions of the child, written to accompany the ratings, follow.

Age. 4-7
Rater: D_1

> Usually gives the appearance of being neat and tidy Her skin is lovely and clear and her eyes a beautiful deep blue. She walks with grace To me she is growing more beautiful each year.

Rater: D_2

> *S.* has an infectious smile, with a dimple which makes her more attractive when she laughs and smiles. She has a lovely, smooth, white skin, her hair is very blond and fine Her clothes are much more attractive than formerly and often she and her sister are dressed alike She walks with ease and grace, and can have a soft voice, but on the whole talks in a loud voice

Age: 4-10
Rater: D_1

Has the features for growing into a very good-looking child, though not beautiful now. Her eyes are a gorgeous blue, but her mouth often looks a little pouty, and her voice is babyish. Her clothes always look well

Rater: D_2

S's body proportions are good She seems very attractive when she smiles, because of her dimples and her sparkling eyes Her mother often dresses *S.* and her little sister alike, and they seem to be wearing more attractive clothes than formerly. Her voice is not as soft, because she shouts and calls out to others in a loud voice She is clean on arriving at school, but is not tidy about her clothes, for they become rumpled and dirty very quickly.

Respect for Property Rights.

The child was rated seven times on this scale between the ages of 1 year, 10 months and 4 years, 9 months, but only three times by the constant rater D_1. This graph made from these three ratings shows fairly close agreement between the raters. There is a slight trend downward, but this variation is probably insignificant. The first rating of the child indicates that she had a fairly high respect for property rights, and she maintained this level. The teachers' descriptions corroborate the ratings.

Age: 4-5
Rater: D_1

Quite attached to her own things, and yet not selfish with them or unpleasant if other children use them. Ideas of property rights very mature and she holds to them well and expects others to do so

Age: 4-9
Rater: D_1

Very mature in regard to property rights. Respectful and would like to see others the same.

Rater: D_2

S has a keen sense of property rights, for she often attempts to settle disputes about toys belonging to others, or tells an adult about it She will take equipment from others for the purpose of giving it to the owner School toys she takes from others if she wishes to play with them She is very careless of her own possessions and never can remember where she leaves things

Response to Authority

On the *Response to Authority* scale, the ratings show considerable variation from time to time. Children seem to respond differently to different persons, and since this scale is based on adult-child relations this difference in responsiveness shows up readily

The rating at 2 years, 5 months is an excellent example of this difference of opinion about a child D_1 was the head teacher, had known the child longer, and had had more experience with children, but *B* was the child's favorite teacher. The two raters gave opposite pictures of the child The only statement upon which they agreed was that the child experimented with new authority to see how far she could go According to D_1 the child was resistant to suggestion, contemplated legitimate suggestion a long time before acting upon it, and then said *"No"* but complied; she would frown, shrug her shoulders, and pout or stamp her foot when a suggestion was made, and pretend not to hear. She was defiant of authority, and resisted a suggestion if it was not about things she herself had planned On the other hand, *B* pictured the child as responding without undue delay to authority, and as resistant only when in a particular mood On the whole, *B* thought her cooperative and responsible

This is an interesting example of a fairly frequent variation in the use of the scales. Either the child's behavior is different in response to different people, or they interpret the same behavior differently

The graph shows that the child's lowest score on this scale was at 3 years, 9 months, when she was rated by *T*, who said she cried if she had to submit to authority, and rebelled by temper-tantrums, hitting, and kicking *D* also scored her low, but not quite so low as *T*. Although her final rating is higher, she never during her stay in nursery school developed into a child who responded well to authority. Her percentile score was always below the median and most of the time below the twenty-fifth percentile The teachers' descriptions bear out the impression one gets from the ratings, that she responds to authority with reservations.

Age: 4-5
Rater: D_1

Her response to authority depends upon three things. Her respect for the authority, whether or not she wants to do the tasks that are requested of her; the mood that she is in. She is much more likely to be in conflict with students than teachers.

When *S.* is agreeable she is constructive and suggestible. *S* generally has to be given time to consider suggestions before she responds.

Rater: D_2

S complies willingly with requests and seldom dawdles as she used to do. However, she is not responsible and dependable, for she often forgets what was asked of her. She never can remember where she left such things as hair ribbons, pajamas, or toys. She is resistant only if in a silly mood and is so carried away by the silliness that she needs reminding to fit into routine required at that time. There is no rebellion or defiance of authority, as there used to be.

Age: 4-9
Rater: D_1

I find *S.* very difficult to check. If she respects the authority she complies nicely, if she is not sure, she behaves badly to discover what may happen, unless she is intensely interested in what she is doing. She can be and is cooperative and responsible in some things and when she desires to be.

Rater: D_2

S enjoys experimenting with new authority and is likely to be silly or disobedient with new students. On the whole, whether she responds willingly to authority depends upon her mood. She has her own ideas and quite often will resist if suggestions are made to her which do not fit in with her own ideas. She is likely to be unreliable and if she thinks the teacher is not looking will become noisy and silly, though she knows it is not the behavior expected.

Sociability with Other Children.

With the exception of one rating when the child was 2 years, 4 months old, the scores on this scale were low at first but later showed considerable improvement. This trend is shown in the statements checked, those up to the age of 3 years, 8 months indicating that she paid little attention to other children and found it difficult to make friends with them After that age she was more pleasant with other children, paid more attention to them, and was able to secure better cooperation from them She had particular friends with whom she was intimate and was unhappy if she was not playing with them The descriptions written by the teachers to accompany the last ratings, before she left the nursery school, follow

Age 4-8
Rater. D_1

> At the age when a child usually pairs off with another child, she is almost always playing in combination with one or more Good balance between leadership and followership. Generally her leadership is purposeful and constructive, though she can become very "giggly" and silly when others pay attention

Rater D_2

> *S.* makes friends with both older and younger children However, she does not mind playing alone She enjoys helping younger children, especially her younger sister, whom she waits on and tells what to do next Sometimes *S* is moody and will quarrel with others over her possessions or small events. She is critical of others and wants everyone to do the right thing and often criticizes the children or tells a teacher. She enjoys conversation while working at a table with other children.

Tendency to Face Reality

Throughout the nursery-school ratings on this scale there was disagreement Rater D_1 rated the child consistently lower than the three assistant raters *G*, *B*, and D_2 The ratings at 4 years, 10 months of D_1 and D_2 are in fairly close agreement The first ratings made after the child's entrance to the recreational club also show very close agreement It is interesting to note that the percentile score decreased steadily from 2 years, 6 months to 4 years, 10 months After she had been in the recreational club for about three months her rating was in the eightieth percentile The highest previous percentile rating was 50

The greatest disagreement occurred at 2 years, 10 months, when *B*, who was the child's favorite teacher, gave the child a score of 85 and D_1 a score of 46. The two raters agreed on but one statement, that the child had a wide range of constructive interests *B*, who scored the child high, said that she faced the issue squarely, concentrated her energy to accomplish a difficult task, did not worry over trivial matters, accepted necessary facts as a matter of course, recognized and accepted the superiority of another child, and accepted the fact that she could not excel in every type of activity; that she appreciated and adjusted to a good authority, accepted just criticism willingly, and did not make excuses for her shortcomings; that she did not lose a sense of reality even in imaginative play. On the other hand, D_1 found that she dawdled to avoid a difficult task, felt unduly disappointed when her plans went wrong, refused to accept certain necessary facts, and found it difficult to accept her own limitations or to accept just blame for her faults; that she regressed to babyish behavior in the face of difficulty, and though she knew when she had failed at a task, became antagonistic in the face of it.

At the ages of 4 years, 6 months and 4 years, 10 months, the teachers wrote descriptive statements to accompany the ratings.

Age· 4-6
Rater: D_1

> Not very willing to accept difficult or hard tasks She makes a pretense of trying and generally day-dreams So far she has not learned to take criticism; she excuses herself, even if it is a simple matter of taking off her hair ribbon and leaving it somewhere. She will slip outdoors without all her wraps, though she knows it means that she must go in Sometimes will fib a little to avoid the issue.

Rater: D_2

> *S.* will work for a long time on a project and is very proud when she succeeds in completing it and will show it to everyone. She knows when she doesn't succeed; she will tell you she couldn't do it, after she has attempted it. She can quickly jump from imaginative play and become matter of fact. However, she will not accept necessary facts until she has asked why a thing should be carried out. She is very critical of her own work and that of others and is willing to be criticized.

Age: 4-10
Rater: D_1

> Continues at times to get into conflict with students because she thinks she can, and she hates to take suggestions or com-

> mands from anyone except herself However, she is really accepting more responsibility for her own behavior recently

Rater D_2

> *S.* dawdles if a task is hard or not to her liking She accepts the fact that others are able to do things as well as herself, but knows when she has done well or has not succeeded She is much pleased when she has succeeded. She will pout or act babyish if the task is hard, or if she is blamed for her faults When accused, she usually has an excuse for her behavior If not accepted by a group, she becomes defiant and tries to get others to follow her in some other activity

It will be noted that D_1 said the child had not learned to take criticism, while D_2 said she was very critical of her own work and that of others and was willing to be criticized This contradiction in opinions shows that the lack of agreement in ratings is not to be attributed entirely to a faulty rating scale, but is due at least in part to something more fundamental—the fact that judgments of other people's behavior are based upon limited experience and upon the reciprocal relation between the rater and the rated

PERCENTILES

Percentiles have been computed for each of the scales Separate percentiles were first worked out for each age group, but wherever age differences were lacking, age groups were combined

The percentiles are based on scores of Merrill-Palmer nursery school and recreational club children who were rated by the teachers and club leaders Since these children are superior to the general population in intelligence and home background, the percentiles are not applicable to children in general, but should be suitable for use with children who are superior to the general population in intelligence and socio-economic status It would be interesting to use the scales with *WPA* nursery-school children and compare the results with those obtained on the Merrill-Palmer children.

Percentiles are based on the average of two raters' scores. Since there is likely to be considerable variation in the scores, it seemed that averaging the two scores would give a truer picture of the child's development All studies of rating scales show that the use of several raters makes for greater reliability. But to make the scales practicable, we have used only two, since in most nursery schools there are not more than two raters who know a child well enough to judge him with any degree of accuracy. Table 34 shows the number of ratings used at each age level to compute the percentiles

TABLE 34
NUMBER AVERAGE RATINGS AT EACH AGE LEVEL USED FOR PERCENTILES FOR NINE RATING SCALES

Age in months	No average ratings	Age in months	No average ratings
Ascendance-submission		Respect for property rights	
24- 47	209	24- 35	72
48-143	288	36- 47	114
144-203	33	48- 59	121
Attractiveness of personality		60-155	96
24-203	540	Response to authority	
Compliance with routine		24-203	604
24- 47	135	Sociability with other children	
48- 59	101	24- 35	91
Independence of adult affection		36- 47	142
24-203	575	48- 59	146
Physical attractiveness		60- 83	58
24- 59	334	84-107	58
60-203	210	108-131	53
		132-203	56
		Tendency to face reality	
		24-191	549

Percentiles are presented in Table 35. For all the scales except *Compliance with Routine,* percentiles are given for children beyond nursery-school age. As yet there are no data available for continuous ratings of children throughout the age range, but a group of former nursery-school children now in the recreational clubs are being rated twice a year on all scales applicable to older children. Theoretically 60 is the median scale value, but for most of the scales the median score is in the seventies, indicating a high level of response for the children whose scores were used for the percentiles

DIRECTIONS FOR USING THE SCALES

The scales were devised primarily for use in nursery schools, but have been used to some extent with school-age children. It should be kept in mind, however, that their reliability has not been established for older children, nor has their adequacy as a measure of personality development been studied sufficiently.

TABLE 35

PERCENTILE RANKS IN TERMS OF SCORE VALUES

Chronological age in months	Percentiles												
	1	5	10	20	30	40	50	60	70	80	90	95	99
Ascendance-submission													
24- 47	29	36	41	49	54	57	60	63	67	70	75	80	84
48-143	32	36	41	50	56	61	66	71	75	78	83	86	94
144-203	31	36	43	54	62	67	71	75	78	82	86	88	90
Attractiveness of personality													
24-203	35	48	53	61	65	69	72	75	78	80	84	87	90
Compliance with routine													
24- 47	36	47	54	62	68	72	77	80	83	87	90	93	95
48- 59	37	46	53	65	74	78	81	83	86	89	92	93	95
Independence of adult affection and attention													
24-203	29	40	46	54	59	64	67	71	74	78	82	85	90
Physical attractiveness													
24- 59	48	58	63	67	70	73	75	77	79	82	85	87	90
60-203	40	49	54	60	64	67	70	72	75	78	81	84	89
Respect for property rights													
24- 35	27	33	39	51	61	67	73	78	82	85	87	89	90
36- 47	14*	34	45	56	63	70	76	80	83	86	90	92	95
48- 59	28	38	49	64	70	75	78	81	84	86	88	89	90
60-155	32	44	54	62	69	75	80	82	85	88	91	93	95
Response to authority													
24-203	8	13	16	20	24	27	30	32	34	37	38	39	40
Sociability with other children													
24- 35	27	34	38	46	53	58	65	70	76	80	86	90	94
36- 47	32	39	45	52	58	64	70	75	80	84	88	91	94
48- 59	36	44	49	57	63	68	73	77	80	84	88	91	95
60- 83	42	46	51	59	64	69	74	78	81	85	89	92	95
84-107	39	52	56	62	66	69	73	77	81	85	89	91	95
108-131	42	49	56	62	67	71	76	78	82	86	90	92	95
132-203	47	53	59	66	71	76	79	82	84	87	90	92	95
Tendency to face reality													
24-191	32	44	50	57	63	68	72	76	80	83	87	90	94

*Percentile influenced by two unusual cases

Directions for rating are printed on each scale, but a few supplementary directions should be kept in mind. It is desirable to have at least two raters and to use the average of the two ratings as the child's score. The raters should know the child well in approximately the same situations. The raters should base their judgments on actual behavior rather than on general impressions In interpreting any differences that may occur in the ratings it is helpful to have the teachers write descriptions of the child's behavior on the back of the sheet and to record any incidents which have influenced their ratings of the child. A child should be rated in comparison with other children of the same chronological age, for it would not be fair to rate the sociability of a two-year-old child on the same basis as that of a five-year-old child. The rater should read all the items of the scale carefully but should check only those which pertain to the child being rated Ratings by different people should be made within the same week in order that judgments may be made on approximately the same behavior.

Scoring the scales is a relatively simple process. The printed key is placed at the left of the column of parentheses in which the rater checks the statements. The scale value of each statement checked is written in the margin The score is found by adding the scale values of the items checked and dividing the total by the number of items checked. Scores for the two raters should be averaged and a percentile rank can then be read from Table 35 If the score falls between the values given in the table, interpolated percentile values should be used.

Graphs of the scores at different times on the nine scales make interesting pictures of a child's personality development. The variations often found between the two raters serve as a basis for a discussion of the reasons for the differences in rating. Changes in the direction of the curve may be studied in relation to other factors in the child's development which may have some bearing on his behavior as observed in the school. By plotting the percentile scores for the nine scales on the same graph the relation between the various aspects of personality measured may be studied.

SUMMARY

By means of a series of rating scales developed according to Thurstone's method for the measurement of attitudes, the following nine aspects of personality were studied: *Ascendance-Submission, Attractiveness of Personality, Compliance with Routine, Independence of Adult Affection and Attention, Physical Attractiveness, Respect for Property Rights, Response to Authority, Sociability with Other Children,* and *Tendency to Face Reality.* Percentiles are given for each of the scales. The following results were obtained.

1 Reliability for the nine scales varied from scale to scale but in general was as high as is usually expected for this type of data.

2 With the exception of sociability and physical attractiveness, age and the personality characteristics measured by these scales showed very little relation

3 Parents rated their children higher than did the teachers on the scales *Ascendance-Submission* and *Attractiveness of Personality;* lower on *Response to Authority, Independence of Adult Affection and Attention,* and *Tendency to Face Reality,* and the same on *Physical Attractiveness.*

4 Intercorrelations for the nine scales showed that the scale *Ascendance-Submission* showed practically no relation to any of the others. Three scales, *Response to Authority, Compliance with Routine,* and *Respect for Property Rights,* were closely enough related to suggest that they might be combined in some way The scales *Attractiveness of Personality* and *Sociability* show sufficiently high relation to indicate a common factor

5. Ratings on the scale *Attractiveness of Personality* (involving presumably subjective evaluations) show greater reliability than ratings on the scale *Physical Attractiveness* (involving presumably more objective evaluations)

6. Several of the scales are found to have too few statements in the middle sections, so that the scales tend to break in two.

7. Some of the scales show insufficient range in statements, restricting the possible uses.

8. Two forms have been prepared for several of the scales The shorter one shows sufficient reliability, as compared with the original longer one, to allow substitution of the shorter form.

9 Where two forms of a scale were used, one with statements in order of scale values and the other in some other logical arrangement, the form giving statements in order of scale values proved to be the less satisfactory

10 Variability characterizes these personality ratings and confirms our opinion that personality is dynamic rather than static.

11 For most of the scales the fifty percentile score is higher than would be expected, indicating a high level of response in the children used in establishing the percentiles.

12. The repeated ratings of children over a period of years show great variation and indicate considerable fluctuation in personality

development, at least as manifested to those who work with the children and observe them daily. Such findings point to the need for long-time evaluations of other rating methods to determine whether the fluctuations are the result of actual changes in the child or are due to the inadequacy of the method.

13. Contradictory opinions of a child's characteristic behavior, as expressed in the descriptions written by the teachers to accompany their ratings, indicate that lack of agreement in ratings cannot be accounted for entirely by a possible faultiness in the rating scale, but must be attributed partly to something more fundamental—the fact that judgments of other people's behavior are based upon limited experience and upon the reciprocal relation between the rater and the rated.

References

1. Atkinson, F. C. Measuring independence in the young child. Merrill-Palmer School, Unpublished Study, 1932. Pp. 62.
2. Bickham, E. The development of attitudes toward property in young children. Merrill-Palmer School, Unpublished Study, 1932. Pp. 63.
3. Bowley, A. H. A sociability study of young children. Merrill-Palmer School, Unpublished Study, 1932. Pp. 54.
4. Conrad, H. S. A statistical study of ratings on the California behavior inventory for nursery school children. *Genet. Psychol. Monog.*, 1934, **16**, 1-78.
5. Davis, F. C. An abbreviation of the scale for rating the tendency to face reality. Merrill-Palmer School, Unpublished Study, 1931. Pp. 6.
6. Ebsley, G. A study of the effect of the sequence of statements on the rating of preschool children on compliance with routine. Merrill-Palmer School, Unpublished Study, 1933. Pp. 14.
7. ———. A study of teachers' ratings of the attractiveness of personality and physical attractiveness of nursery school children. Merrill-Palmer School, Unpublished Study, 1933. (Accepted as a thesis for the degree of Master of Arts, Wayne University.)
8. Gerlaugh, M. E. An evaluation of personality traits of preschool children by their parents. Merrill-Palmer School, Unpublished Study, 1932. Pp. 84. (Accepted as a thesis for the degree of Master of Arts, Ohio State University.)
9. May, M. A. The foundations of personality. (In) Achilles, P. S. (Ed.): *Psychology at work.* New York: Whittlesey House, 1932. Pp. xiii+260.
10. McClure, T. F. Factors affecting the response of preschool children to authority. Merrill-Palmer School, Unpublished Study, 1931. Pp. 50.
11. Reynolds, W. R. Rating of a group of preschool and preadolescent children on their tendency to face reality. Merrill-Palmer School,

Unpublished Study, 1934 Pp 56 (Accepted as a thesis for the degree of Master of Arts, Ohio State University.)

12 STRAYER, M J. Age factors in sociability of children. Merrill-Palmer School, Unpublished Study, 1931. Pp 26. (Enlarged and accepted as a thesis for the degree of Master of Arts, Ohio State University, 1932)

13. SYMONDS, P M Diagnosing personality and conduct. New York Century, 1931 Pp xvi+602

14 THURSTONE, L L, & CHAVE, E J The measurement of attitude Chicago Univ Chicago Press, 1929 Pp vii+96

Merrill-Palmer School[4]
Detroit, Michigan

---

[4] *Acknowledgments* This study represents the cooperative thinking and working of many of the staff and students of the Merrill-Palmer School. Miss Leone Chesne, statistician at the School from 1928 to 1930, was responsible for the scaling of the ratings, and since leaving has given generously of her time both in actual statistical work and in an advisory capacity Dr. L L Thurstone of the University of Chicago advised with us in regard to scaling the statements and the elimination of irrelevant ones. We are grateful to the large group of staff and students who sorted the statements preparatory to the formation of the scales. To the raters—Mrs Florence Willson Duhn, Miss Ruth Dales, Mrs. Elise Hatt Campbell, Miss Joan Millington, Miss Julia Dawson, Miss Winifred Harley, Miss Elna Jensen, and others too numerous to mention individually—who over a period of years have painstakingly rated the children, we express our appreciation. Without their willing and intelligent assistance the study would not have been possible The students who made special studies using the scales are mentioned by name and their studies are listed in the references We thank them for their contributions We are indebted to Mr. Wendell Vreeland of Wayne University for advice in regard to the computation of the percentiles To Dr E. Lee Vincent who made helpful criticisms and suggestions, and to Miss Dorothy Tyler who edited the manuscript, we express our gratitude.

KATHERINE ELLIOTT ROBERTS AND RACHEL STUTSMAN BALL.

The Journal of Genetic Psychology, 1938, 52, 151-157

THE RELATIONSHIP BETWEEN THE SUPERIOR COLLICULUS AND THE STRIATE AREA IN BRIGHTNESS DISCRIMINATION*

Department of Psychology, Harvard University

E. E. Ghiselli[1]

The fact that the brightness discrimination habit can be formed in the absence of the striate area has led to the suggestion that the habit might be mediated by the superior colliculus. However, recent evidence has shown that rats with lesions in the superior colliculus form the habit as readily as do normal animals (1, 7), and that such lesions produce no loss in the retention of the habit or change the threshold of brightness discrimination (2). On the basis of these findings it has been pointed out that there might be an equivalence of function between the striate area and the superior colliculus (2, 6).

Inasmuch as cases with lesions both in the striate area and the superior colliculus might throw further light on this matter, the following experiments were undertaken.

Methods and Procedure

Nine hooded rats were used as subjects. The apparatus for measuring brightness discrimination was the modification of the Yerkes box devised by Lashley (6) Food was used as the incentive to learning, and electrical shock was administered as punishment for incorrect responses In both learning and retention tests 20 successive errorless trials were taken as the criterion of successful performance.

Lesions were produced in the superior colliculi of all cases. Following a 15-day recovery period the animals were trained in the discrimination box. After a 10-day period they were tested for retention, retrained to the criterion, and subjected to lesions in the striate area. After another 10-day recovery period the animals were again retrained on the brightness discrimination.

---

*Received in the Editorial Office on May 5, 1937.

[1] National Research Council Fellow The writer wishes to express his thanks to Professor K S Lashley for criticism and advice

In performing the operations sodium amytal and ether were used as anasthetics. The methods employed in producing the subcortical lesions have been described elsewhere (2). A thermo-cautery was used in the production of the cortical lesions.

At the completion of the experiment the brains were removed from the animals, fixed in alcohol, and mounted in celloidin. The brains were sectioned at 50mu. and stained with thionin. Camera lucida drawings were made of the stratum opticum of the superior colliculus and the extent of destruction in that layer estimated. The cortical lesions were reconstructed, and the amount of destruction measured, after the method of Lashley (4). Inasmuch as destruction of the striate area produces retrograde degeneration in the lateral geniculate body pars dorsalis (5), the extent of degeneration in this nucleus was also determined.

RESULTS

The per cent of destruction of the stratum opticum of the superior colliculus of each animal will be found in Table 1. In this table

TABLE 1

DESCRIPTION OF LESIONS AND TRAINING RECORDS OF INDIVIDUAL ANIMALS

No.	Per cent colliculus destruction *L*	*R*	Per cent cortical destruction	Learning Trials	Errors	Relearning Trials	Errors	Relearning after striate destruction Trials	Errors
1	P 30	P 20	23	15	8	17	1	34	10
2	P 80	PM 80	28	18	6	4	1	26	11
3	PM 60	P 40	28	30	5	0	0	25	5
4	PM 50	PM 50	29	28	7	1	1	40	10
5	P 70	PM 40	33	24	5	0	0	38	13
6	PM 50	PM 50	17	69	16	0	0	0	0
7	PM 70	PM 70	21	31	11	0	0	28	11
8	PM 60	PM 60	28	63	16	2	1	27	9
9	PM 50	PM 20		55	16	0	0		

P and *M* refer respectively to the posterior and medial portions of the stratum opticum. In every case the subcortical lesions were confined to the superior colliculi, but in no instance was any layer of the colliculus completely destroyed. The reconstructions of the cortical lesions will be found in Figure 1. Three of these animals,

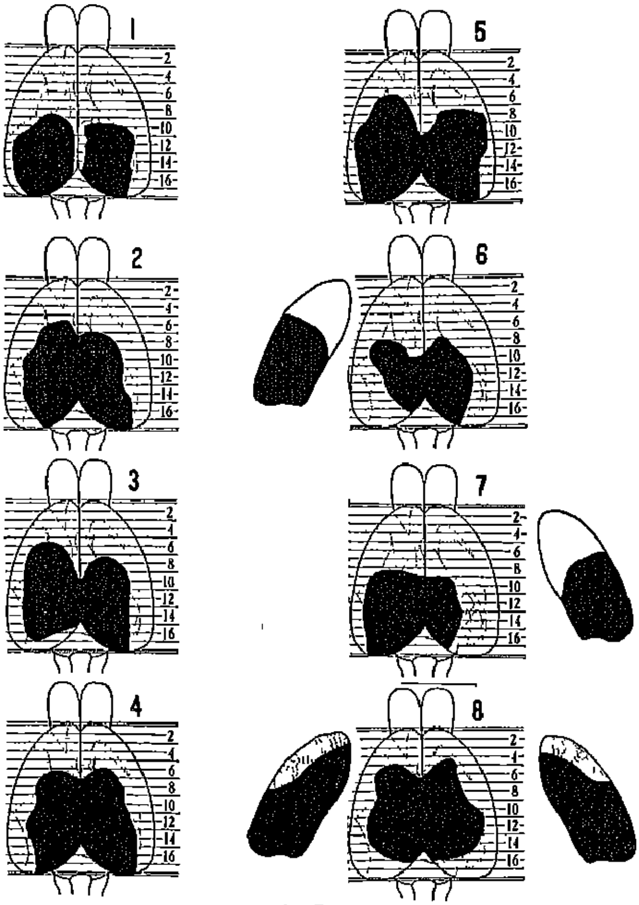


FIGURE 1
Reconstructions of Cortical Lesions

rats 6, 7, and 8, had incomplete destruction of the striate areas. For these cases a section through the lateral geniculate bodies which retain undegenerated portions is shown accompanying the diagram of the cortical lesion. The black portions indicate the regions with complete retrograde degeneration and the stippled portions regions in which only a scattering of normal cells appear. The per cent of cortical destruction is also given in Table 1.

The individual training records of the animals on the various tests are presented in Table 1 In Table 2 are to be found the mean number of trials and errors involved in the learning and relearning tests. In both tables the figures for trials exclude the 20 criterion trials.

In Table 2 the initial learning of the animals with colliculus lesions is compared with that of a group of normal animals. These normal animals had had previous training in the discrimination of black and white cards and striations on the jumping stand (2). In spite of the previous training on the part of the normal rats, the operated animals were only slightly inferior to them. The differences in both trials and errors are insignificant, the critical ratio of the former being 59, and of the latter 96. Furthermore, the rats with colliculus lesions fall well within the range of practice required by the normal animals. The range of trials of the normal rats was 2 to 104, and of the operated animals 16 to 59; the range of errors made by the normal animals was 1 to 25, the range of the operated being 5 to 16.

TABLE 2

AVERAGE PERFORMANCE ON VARIOUS TESTS

			N	Trials	Errors
Initial learning-normal			25	32.60	8.20
Cases with Colliculus Lesions	Initial learning		9	39.56	10.00
	Retention		9	2.89	0.44
	Retention after striate destruction	All cases	8	27.25	8.63
		Cases with complete striate destruction	5	32.60	9.80
		Cases with partial striate destruction	3	18.33	6.67

The correlation between the amount of practice required to form the brightness discrimination habit and the extent of destruction in the stratum opticum of the superior colliculus fails to reveal any suggestion of a relationship between these two variables. The rank order correlation between size of lesion and trials was .08, the correlation with errors being —.08.

Thus, these results confirm those of Freeman and Papez (1), and Layman (7), which indicate that lesions in the superior colliculus have no effect upon the rate at which the brightness discrimination habit is formed.

Lashley (3, 6) has found that after destruction of the striate area the brightness discrimination habit is lost, but may be relearned He found correlations of the order of 60 to .70 between the extent of cortical destruction and the rate of relearning The same results were found in the present experiment for the animals which learned the habit with colliculus lesions and were tested for the retention of the habit after destruction of the striate area. It will be observed from Table 2 that the retention after striate destruction was markedly poorer than was the pre-operative retention The rank order correlation between the extent of cortical destruction and the number of trials required in relearning after striate injury was 60, the correlation with errors being .62. Apparently, then, the relationship between the extent of cortical destruction and the amount of loss of the habit is unchanged despite the fact that the habit was formed with only part of the superior colliculus functional.

Lashley (6) has also shown that the above mentioned relationship is not indicative of a continuous function between the extent of cortical destruction and the degree of amnesia Rather the post-operative loss of the brightness discrimination habit follows an all or none principle Complete loss occurs after complete destruction of the striate area, and perfect retention may occur if only a small portion of the striate area remains intact. In three of the animals in the present experiment, rats 6, 7, and 8, the striate area was incompletely destroyed. All of these animals made better records in relearning after striate destruction than they did in initial learning. One of these, rat 6, made a perfect score in retention although the visual area was completely destroyed on one side and half on the other. The relearning scores of the five animals with complete destruction of the striate area differ but slightly from the initial learning scores of the whole group (Table 2). Thus, animals which learn a brightness discrimination with lesions in the superior colliculus behave like animals with that nucleus intact in the retention of the problem after destruction of the striate areas.

Discussion

The results of the present experiment add to the growing mass of evidence (1, 2, 7) indicating that the brightness discrimination habit is unaffected by lesions in the superior colliculus. Lashley's

(4) results show that while destruction anywhere in the cortex does not retard the learning of a brightness discrimination, when direct injury is produced in the subcortex in addition to the cortical destruction, learning is retarded. A similar situation obtains in the relearning after removal of the striate areas. When the lesion is confined to the cortex the habit is relearned in normal time. However, when there is direct subcortical destruction in addition to the cortical injury, relearning is done at a slower rate (6) It was impossible for Lashley to judge from his data whether to ascribe the retarded learning to destruction of the superior colliculus or of the subcortical structures lying just anterior to it. In the first place he had no cases where the subcortical involvement was clearly limited to the superior colliculus, and in the second place in those cases with destruction in the pretectal area, the lesion interrupted the optic fibers to the colliculus. The data of the present experiment shed some light on this problem. The subcortical lesions of the cases herein reported were completely confined to the superior colliculus, and even in the cases with the complete destruction of the striate area there was no indication of a retardation in relearning. Thus, the retardation of Lashley's cases cannot be attributed to destruction of the superior colliculus alone. However, it is not yet clear whether the retardation is to be ascribed to destruction of the nuclei of the pretectal area or to destruction in all of the primary subcortical visual nuclei.

SUMMARY

1. A group of rats with lesions in the superior colliculus were taught a brightness discrimination, and after the habit was established the striate areas were destroyed.

2. It was found that the colliculus lesions did not affect the rate of learning, and that after complete destruction of the striate areas the habit was lost but was relearned in normal time, while after incomplete striate destruction the retention might be perfect.

REFERENCES

1 FREEMAN, G. L., & PAPEZ, J. W. The effect of subcortical lesions in the visual discrimination of rats. *J. Comp. Psychol.*, 1930, **11**, 185-192.

2 GHISELLI, E. E. The superior colliculus in vision. (In press)

3. Lashley, K S. Studies of cerebral function in learning. VII The relation between cerebral mass, learning and retention *J. Comp Neur*, 1926, **41**, 1-58
4. ——— Brain mechanisms and intelligence Chicago. Univ. Chicago Press, 1929
5. ——— The mechanism of vision VII The projection of the retina upon the cerebral cortex of the rat *J Comp Neur*, 1934, **60**, 57-79
6. ——— The mechanism of vision. XII Nervous structures concerned in the acquisition and retention of habits based on reactions to light *Comp. Psychol Monog.*, 1935, **11**, 43-79
7. Layman, J. D Functions of the superior colliculi in vision *J. Genet Psychol*, 1936, **49**, 33-47.

Department of Psychology
University of Maryland
College Park, Maryland

The Journal of Genetic Psychology, 1938, 52, 159-163.

LEARNING WHICH MODIFIES AN ANIMAL'S SUBSEQUENT CAPACITY FOR LEARNING*

Department of Psychology, University of Rochester

LEONARD CARMICHAEL[1]

An elevated maze shown in Figure 1 was set up in a room in which indirect illumination was constant and in which temperature and sound were also both relatively constant

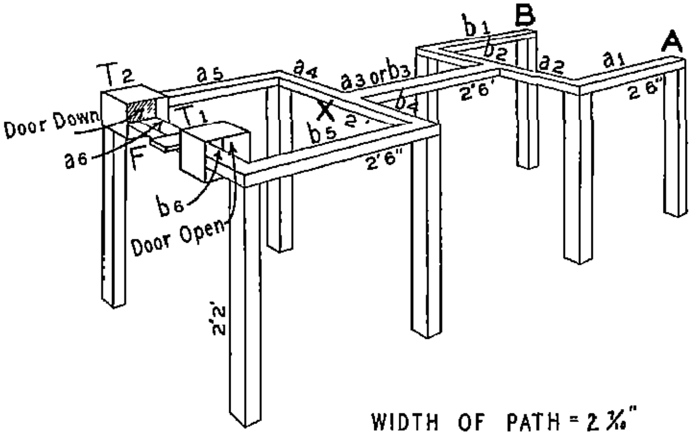


FIGURE 1

The maze is similar to one previously described by Carmichael and Marks in an experiment using cats as subjects (1) and to one described by Trueblood in an experiment using rats as subjects (2)

In the present experiment, 10 young but mature male rats of Wistar stock were trained. Ordinary precautions concerning diet and care of health were adhered to, and weight and general physical condition of the animals seemed as satisfactory at the conclusions of the experiment as at its beginning

---

*Received in the Editorial Office on June 18, 1937

[1] The writer wishes to express his gratitude to Mr. H A Israel who assisted in these experiments which were conducted in the laboratory of psychology of Brown University

The animals were placed in regular alternation at starting points *"A"* and *"B."* If placed at *"A,"* they ran a maze *a*1, *a*2, *a*3, *a*4, *a*5, and *a*6, and received food. If placed at *"B,"* they ran *b*1, *b*2, *b*3, *b*4, *b*5, and *b*6, and received food. Food at *"F"* was at all times clearly in view. The difference in the maze was merely determined by opening and closing a hidden door in one or the other of the tunnels, *T*1 or *T*2. Care was taken to avoid differential secondary clues concerned in door manipulation.

A learning curve, plotted in terms of errors, for this regular alternation habit is shown as Curve I in Figure 2 and Table 1. As this curve shows, virtually complete mastery had been secured by the sixteenth trial. One trial consisting of 12 runs was given at the same time each night. Practically the only errors made were at choice point *"X,"* and these only were counted.

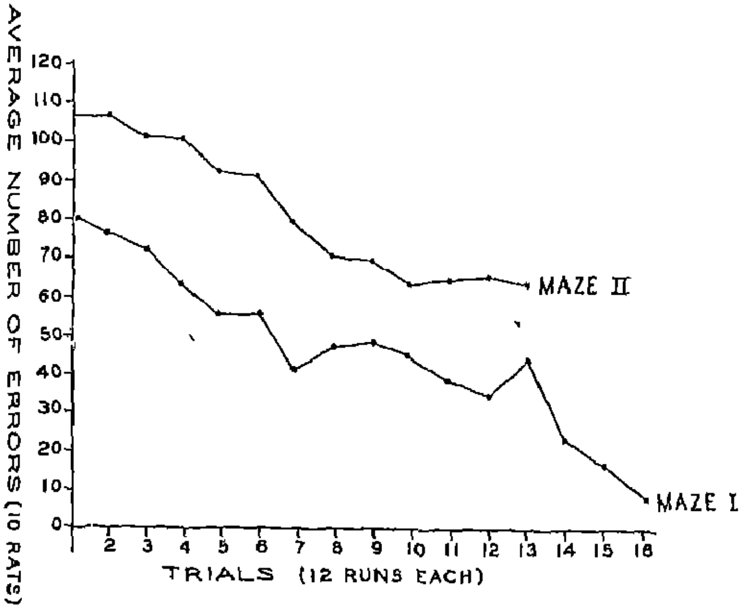


FIGURE 2

After the first habit had been mastered, the same animals were each given 36 runs with irregular alternation (that is, so-called coin-tossing alternation) on the same maze pattern. They showed

TABLE 1

Error Record—Maze I

Trial no (12 runs each)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Av no errors (10 rats)	80	76	72	63	55	55	41	47	48	45	39	35	44	23	17	9

definite indications of learning this habit, but this part of the experiment is not relevant to the main part of the problem here under consideration

The animals were now placed upon the same maze in regular alternation at starting points "*A*" and "*B*," but the doors which were closed were the opposite of those which had been closed in the preceding part of the experiment Thus, the animals were now required to learn that if placed at "*A*," the correct run was *a*1, *a*2, *a*3, *b*4, *b*5, and *b*6, or if placed at "*B*," the run was *b*1, *b*2, *b*3, *a*4, *a*5, and *a*6.

The results of this new training situation are shown in Curve II in Figure 2 and in Table 2. It will be noticed that at the conclusion of 156 runs on this second maze, given in a series of 12 each on 13 consecutive nights, the animals had not learned the new maze and had indeed not reached the level of performance which they had reached on the first maze after 48 runs on four consecutive nights

TABLE 2

Error Record—Maze II

Trial no. (12 runs each)	1	2	3	4	5	6	7	8	9	10	11	12	13
Av no errors (10 rats)	106	106	101	100	92	91	79	70	69	63	64	65	63

Time curves and modified Vincent curves have been calculated for this same data, but the facts indicated by these calculations are not especially significant.

Qualitatively, the behavior of the animals on the second form of the maze was remarkably uniform When the animals that had been trained on Maze I had been given a few trials on Maze II, their behavior began to change in a striking way. The choice point marked "*X*" came to lead frequently to delay. Often the animal would stop and turn its head to right and left a number of times before stepping onto *a*4 or *b*4. In certain instances, animals that had previously run the entire maze in an average of not more than five seconds for a trial period of 12 runs would now sit at this choice point in extreme cases for a little over 10 minutes as timed by a stop-watch. This behavior is interesting because extrinsic punishment was never used in the experiment. One door or the other as previously described was always closed, and the total time required to run the maze, even when one door was closed, was not necessarily more than seven seconds. In spite of this fact, the new maze seemed so to interfere with previous learning and previous freedom of movement that, to speak in terms of analogies, the whole behavioral activity of the animal sometimes seemed to be inhibited, and sometimes it seemed to release a new and apparently disorganized set of responses. Thus, a far greater delay frequently resulted than that which would have been required even if the animals had made nothing but error-runs but had kept running at the previous rate. In a few instances, after an error-run, the animals even re-traced back to the starting-point, behavior which had not been demonstrated in the preceding 140 runs In such cases, the animal's general behavior as indicated by head-turning, irregular movement, and the like, seemed typical of the behavior described by those who have set up so-called experimental neuroses in animals. As noted above, it is striking that this should be the case because in this experiment no punishment was used, and the result of an error-run at most delayed the animal in securing food by two or three seconds.

Experiments have shown that in many maze situations, incorrect alleys are often eliminated gradually. In this experiment, on the contrary, if the incorrect path at "*X*" was chosen, there was never any indication, up to the end of the experiment, that the error-run was shortened The direct stimuli received from the closed door in Tunnel *T*1 or *T*2 were apparently always required to act upon the animal before re-tracing took place.

There is little reason to suppose that the second maze was more difficult than the first or that, had the second maze been first presented, it would have required the animals a significantly longer time to earn it than it did the one actually used first. Indeed, experiments by M. O. Wilson on a delayed reaction maze suggest that the second or straight-path maze might have been more readily learned than was the first or crossed-path maze (3, 4)

It thus seems that the present experiment means that stimulus cues which were effective in determining the behavior of animals on the first maze so that learning took place *were closely identical* with those required to set up adaptive behavior on the second maze

It seems that this may be interpreted to mean that at least in this situation, once specific exteroceptive and internal patterns of stimulation have been isolated from a more general pattern and responded to adaptively, a change in the organism takes place. This change is such as to make it less easy for that organism to develop new adaptive responses to a specific new set of stimuli also present in the total situation than it would have been before the previous specific pattern of stimulation had been isolated from the total situation. A theoretical explanation of this procedure in terms of conditioned responses may suggest itself to the reader The facts recorded here once more point, from a slightly novel angle, to the importance of a detailed knowledge of the past of an organism for one who would explain behavior in terms of the dynamic relationships of an organism and its environment as such relationships may be observed at any given moment.

REFERENCES

1. CARMICHAEL, L, & MARKS, L D A study of the learning process in the cat in a maze constructed to require delayed response *J Genet Psychol*, 1932, 40, 206-209
2. TRUEBLOOD, C K, BOUTELLE, W E, & VAUGHN, A E Behavior of white rats in a maze requiring delayed response *J Genet. Psychol*, 1936, 49, 227-231
3. WILSON, M O The delayed response *Psychol Bull*, 1931, 28, 673-674
4. ——— Symbolic behavior in the white rat *J Comp Psychol*, 1934, 18, 29-40, 367-384.

Department of Psychology
University of Rochester
Rochester, New York

The Journal of Genetic Psychology, 1938, 52, 165-181.

FURTHER EVIDENCE OF COLOR DISCRIMINATION IN RODENTS*

Department of Psychology, The University of Nebraska

WILLIAM E WALTON AND RUSSELL W BORNEMEIER

In 1933 the senior writer published the report of a study which gave the first positive evidence of color discrimination in the albino rat In subsequent experiments (10, 6, 4) an error of technique was discovered which rendered certain results equivocal. The values of this experiment are: (*a*) that it reopened the question of color vision in lower mammals, (*b*) that it demonstrated quite satisfactorily that the white rat could discriminate red from darkness.[1]

Two articles had appeared just prior to that of the senior writer, but these cannot be said to have been instrumental in reopening the question of color vision in lower mammals, inasmuch as the results were negative and, therefore, in keeping with those which had preceded them Munn (5) in 1932 published a report in which he had used two animals to set up a discrimination between Hering gray papers and Hering yellow, blue, green, and red. He was fortunate in finding two Hering papers, green and yellow, which were approximately equal in brightness for the rat's eye and approximately Hering gray No 7. He undertook to test color discrimination by training the two animals on these two colors, giving them from 20 to 50 trials a day. If the earlier findings of the writers indicate anything at all, the rats would not be expected to be able to discriminate yellow and green. In Munn's training technique, he

---

*Received in the Editorial Office on July 6, 1937.

[1] Munn (6, p 287 f) seems dissatisfied with the senior writer's statement that tests were made in all of their studies to determine the influence of non-visual cues These included the elimination of red in the red-darkness discrimination study. The writer was not interested at all in determining the stimulus limen, but rather, whether or not the animal could discriminate red from darkness The objection that the exact brightness point at which experimentation was terminated was not given in *cp*'s in no way affects the other findings and is relatively unimportant In the present article, however, these *cp* values are given There is no evidence, whatever, that the results of the Nebraska red-darkness discrimination study are equivocal.

used the same animals on the discrimination problem that had been used in setting up the point of subjective equality (*PSE*). Experience in the Nebraska laboratory has indicated that animals trained upon the brightness problem have considerable difficulty in transferring to a hue problem. In fact, many animals will simply refuse to run when the brightness factor to which they have been responding has been eliminated. Furthermore, his statement (p. 361) that "One is forced to the conclusion, therefore, that the subjects of this experiment were color blind," is not warranted by the data obtained. The two rats were tested only on Hering yellow and Hering green, and such a generalization is not justified.

Many of the recent investigators have cited the work of Coleman and Hamilton (1) as further evidence of color blindness in the white rat. The two experimenters are to be congratulated upon their excellent method. It was their plan to determine the point of subjective equality ("confusion pairs") and then train the same animals at the *PSE* to determine whether or not they could discriminate the colors. Table 1 indicates an error in their technique.

TABLE 1
ILLUSTRATING INCONSISTENCY IN SELECTION OF CONFUSION PAIRS

Confusion pairs			
Blue	Black	Total trials	Per cent correct
36-12A	47-5A	8	25
34-11A	39-6A	11	72
34-11A	38-4A	76	52
33-10A	39-6A	53	58
33-10A	38-4A	13	69
33-10A	46-5A	37	75
33- 8B	46-5A	6	16
33- 7B	46-5A	12*	25*
33- 4B	46-4A	18	55
Green	Black		
25- 7A	46-4A	36*	52*
Red	Black		
2- 3H	39-6A	15	52
2- 3B	38-4A	17*	23*
2- 3B	46-5A	13	47
Red	Green		
2- 1D	25-7A	21	47

*Compare total trials on which percentages were determined and confusion pairs selected

It will be noted that they were quite inconsistent in the number of trials used to determine the "confusion pairs" and, also, in the percentage taken to indicate confusion The trials vary from 6 to 76, and the percentages from 16 to 75. Their next step indicated considerable insight into the problem They selected new animals and attempted to train them on the "confusion pairs." Table 2 shows the total number of trials used to train these new rats. While the percentages are satisfactory, the total number of trials used in each experimental situation is quite inadequate to train rats to discriminate the "confusion pairs" In the Nebraska laboratory at least 800 trials are run in such tests Furthermore, after a so-called failure of the animal to discriminate, they attempted to train these same animals upon combinations in which there were great brightness differences This they were unable to do, but they failed to see the significance of the results on this part of the problem When these same animals failed upon the brightness study, Table 3, it would seem to indicate that learning on the "confusion pairs" had begun and that this learning was interfering with the transfer to

TABLE 2

ATTEMPT TO TRAIN "NEW" RATS ON CONFUSION PAIRS

Confusion pairs		Total trials	Per cent correct
Blue 25-7A*	Black 46-4A	131	55
Green 33-7B*	46-5A	117	58
Red 2-3B*	38-4A	102	45

*See Table 1.

TABLE 3

ATTEMPTS TO TRAIN TO BRIGHTNESS AFTER BEING CONFUSED

Brightness pairs		Total trials	Percent correct
Blue 33-8B	Black "India ink"	161	61
Blue 33-3B	Black "India ink"	20	55
Red 2-3H	Black "India ink"	23	56
Red 1-1B	Black "India ink"	114	64

the brightness habit The inconsistency with which Coleman and Hamilton conducted their tests is indicated by the total number of trials run upon the brightness pairs. In the Nebraska laboratory the writers feel that at least 30 perfect consecutive trials, over two days, should be the criterion of success, followed by 100 trials on each "confusion pair" in which a high percentage of success is indicated. Summarizing the errors of Coleman and Hamilton, the writers find (*a*) inconsistency in determining "confusion pairs" and, therefore, the possibility of the selection of those pairs being unreliable, (*b*) lack of persistence in training the animals and, also, in conducting the control tests, (*c*) failure to observe that the inability of the rats, partially trained on the "confusion pairs," to learn the brightness problem (Table 3) really indicated an attempt to discriminate the colors on the basis of hue differences. The further statement that "One month later the rats were very easily trained, giving nearly perfect scores after 20 to 30 jumps," is additional evidence that training on hue discrimination had been started.

In 1933 the senior writer (9, p 391) had said:

> . . . either these "rod-like" structures which Lashley has found are serving a dual function, or the nerve fibers of the retina are directly stimulated by the light. To verify this contention, we need further studies by physiologists and psycho-biologists We believe, however, that if such an explanation should be found to be the correct one, we will be compelled to modify our present color theories It may be that the much discussed cones in their differentiated functioning merely enrich the color experience.

In concluding the article (p. 394) he stated:

> . . if these conclusions are valid, we may be led to question the assumption that the cones are the organs of color vision, provided, of course, that it can be shown *conclusively* that the retina of the rat's eye is coneless.

In commenting upon this study, Walls, 1934 (8, p. 363) says.

> Walton (1933) has recently reported strong evidence for qualitative hue discrimination in the rat and has attacked the validity of the Duplicity Theory on the ground that the rat "does not have the structures essential to many of the physiological theories of color vision" He cites Lashley (1932) as

> having found rat retina to be pure-rod Walton is led by this to question the assumption that the cones are the organs of color vision.

He follows with evidence that the Albino Norway Rat's retina possesses cones as well as rods and concludes:

> The demonstration of cones in the rat retina is no proof of the color-perceptive function of these elements, but it does render unwise any assumption, such as Walton's, that this animal's rods perform in a dual capacity.

It is questionable whether an assumption was made; rather, a problem was proposed, a question asked to which two answers were given, either one of which might be correct if the previous retinal studies were reliable. The aid of the physiologists and psychobiologists was asked. Walls gives that aid, and his findings supply the missing information needed in the Nebraska studies of the color vision problem Since it has been verified, at least in correspondence, by Lashley, who examined one of Walls' preparations, a stumbling block in the study of color discrimination has been removed It is still possible, however, that the perceptive process is cortical In fact, the expression "color-perceptive function of these elements" is meaningless in the light of modern studies of perception and is undoubtedly based upon a misconception of psychological terms Furthermore, the discrimination of two hues is probably a matter of cortical action, the retinal structures only entering the picture if cones are missing. The question concerning the importance of the cones in color vision remains unanswered; and should the writers be able to show color vision in forms having only one type of nuclei, the Duplicity Theory would be weakened

In 1934 Munn (6) made another attempt to show that rats are color blind In his introduction he mentions his study on green and yellow and again generalizes from the weak evidence presented that "the animals were color blind, a conclusion reached several years earlier by Watson and Watson." It is surprising that in this introduction he supports the work of Coleman and Hamilton[2] The very careful analysis seems to have been upon only those studies which have yielded positive evidence for color vision. The excellent criticism which he makes of the Walton study is in keeping with the results of a series of experiments conducted in the Nebraska laboratory, since the first publication, to determine correct techniques of experimentation.[3] While Munn has discovered some good evidence for equivocality of Walton's results, his attack on the problem of blue-yellow introduced conditions which rendered his results equivocal. The apparatus was built in such a manner that the rats were forced to run across the stimulus patches.[4] This apparatus has been duplicated in the Nebraska laboratory, and the results of studies upon its efficiency have indicated that it is only when there are marked differences between the brightnesses in the stimuli that the animals will learn the problem. At best, the apparatus is a clumsy device and one which makes rapid experimentation impossible. *It is essential in learning problems where two stimuli are difficult to discriminate that the experiment run smoothly, with the minimum lapse of time between trials. At least thirty trials per day are necessary.*

---

[2] A study which has just been shown to be full of errors

The apparatus alone would account for errors in the study. In addition, experimentation was stopped after 500 trials. The writers have often experienced success in discrimination problems only after 800 trials.[5] The more cumbersome the apparatus, the more likelihood that more trials will be required for the learning problem. The controlled experiment, involving a study of lights equated in brightness for the human eye, proved only that the animals could discriminate brightnesses in the apparatus. This is further borne out by the fact that there was no difference between the actions of the trained and untrained animals (6, p. 296). According to the results which have been obtained in the Nebraska laboratory, there should have been a difference between the two groups, *for rats trained upon a hue discrimination have difficulty in transferring to a brightness discrimination*

---

[3] Some of the results of this experimentation have been published See (10).

[4] In the Nebraska apparatus the rats are forced to look directly at the stimuli before making a choice It is significant that in the Nebraska studies there is frequent evidence of the "wavering" response, in which the animal sits in front of the patch and weaves his body back and forth during a critical period in the learning problem. No evidence for this reaction is presented by Munn.

[5] In an exploratory series of experiments, the senior writer has found that a thousand or more trials were necessary on one apparatus, in order that cats and dogs might be trained to discriminate colors which were at the *PSE*.

In both of the articles by Munn, reference is made to the use of wax and other papers as a means of regulating the brightness factor. Experience in the Nebraska laboratory indicates that such a control introduces a "texture" difference between patches to be discriminated The very finest of bond and tissue papers introduce a mottled appearance on the side in which they are used.

Another difference between the experiments involves the time factor Experiments with Munn's apparatus in the writers' laboratory have indicated that it required approximately five times longer to put an animal through one trial on his apparatus than on the Nebraska apparatus. The advantage of using a technique in which the experimenter remains seated at the controls of the apparatus is apparent, for the necessity of carrying the animal from the back of an apparatus around to the front, placing him on the entrance platform, awaiting his response, and then repeating this procedure may act as a serious disturbing factor. *The first task of the subject in any discrimination problem is that of learning what is to be required of him.*

The first positive evidence of color discrimination, other than has been previously cited, is that of Muenzinger and Reynolds, 1936 (4). While their method was somewhat different from the one previously employed in the Nebraska laboratory, it is interesting to note that their rough measurements with the Westinghouse light meter indicated brightness values larger than those of any previous experimenter. In general, they conclude that "the rats in this case present a hierarchy of discrimination habits, consisting of a primary brightness-relation habit, a secondary absolute brightness habit, and a weak color discrimination habit." These were essentially the same conclusions drawn previously by the senior writer (9, p 368)

In the same year (1936) Munn and Collins undertook to check the results of Muenzinger and Reynolds It is significant that the design of the Munn apparatus was radically changed for this study. No reason for this change is advanced. It is obvious, however, to anyone who has used it in the laboratory In the new apparatus the stimuli were placed directly in front of the subjects The experimenters trained some rats to select red (an intermediate brightness) from white and darkness A footnote is appended, however, in which they say, *"The question might arise as to whether red may* *have been darker than both the black and white.*"[6] This extreme skepticism is difficult to understand. If the experimenters use the term black or "darkness" it should mean absolute and not relative black or darkness. The use of bond and wax papers in regulating the brightness factors in certain parts of the experiment is questionable. It is difficult, also, to understand why the distance of the lights to the stimulus patch was not used by the experimenters, since the Munn apparatus was especially designed with that control in mind

After many controls had been run, the authors concluded that "the only constant factor which the animals might discriminate was the wave length property (or properties) of the red light"

Apparently, not willing yet to concede red vision to the rat, a test was conducted in which a study was made to determine whether or not their monochromatic red might not possess some "quality" other than its "redness" to which the rats were responding and which any monochromatic light might possess. To assume that an animal possesses an ability which human subjects with their highly developed visual apparatus do not possess seems a bit far-fetched. However, the negative results are taken by the experimenters to indicate that the rats can discriminate red from darkness and from other colors.

TECHNIQUE AND APPARATUS

The Chromopathometer and discrimination box previously described by Walton (9, p. 375 f)[7] were used in the experiments, with the following modifications: The doors, *d* and *d′* (9, p. 376), were hinged at the back of the box so that they swung outward into the long passageways, *p* and *p′*, thus eliminating the longer path of the animals around the door Small metal hooks were installed to hold the doors in position The hooks were so arranged that they could be released by the action of electro-magnets. Within the entrance to the compartments, *c* and *c′*, were installed platform switches, in such a way that they blended into their surroundings and gave no reflection from the surface. No matter which compartment was entered by the animal, a switch was depressed. It was only upon entering the wrong compartment that the electro-magnets operated to close the door. Thus, the animal automatically punished himself. Small switches at the front of the apparatus enabled the experimenter to regulate this punishment

---

[6] Italics those of the writer's

[7] Walton, W E, & Morrison, B. M. The chromopathometer. *J. Exper. Psychol.*, 1929, **12**, 254-258.

In general, the same methods of training were employed as were used in Walton's earlier experiments (9, p. 383 ff) An "unhesitating" positive, or correct choice, was indicated by means of a plus mark (+), and a definite negative, or incorrect choice, was indicated by means of a minus mark (—) "Wavering" reactions were designated by doubling the marks used above, a "double plus" (╪) if the reaction was correct, a "double minus" (=) if incorrect. By "wavering" reaction the writers are referring to the type of behavior described by Walton (9, p 391), Muenzinger (4, p 204), and Coleman and Hamilton (1, p 178), in which the animals sit in front of the partition separating the two compartments and "weave" their bodies back and forth, as though they were attending to one and then to the other stimulus These reactions will be referred to in a later section of the article

Prior to experimentation, the usual precautions were exercised in preparing the animal for handling The rats were fed by the experimenters "by hand," carried while eating their food, and subjected to such conditioning as would eliminate the fear reactions which usually accompany experimentation upon the discrimination problems Once the rats were used to the human factor, they were placed in the discrimination box and allowed to explore all compartments. Both of the doors, *d* and *d'*, were left open during this preliminary exploration. Then one door was closed, the side being determined by the regular experimental series. Next the lights were introduced and preliminary work begun Finally, punishment was brought in, and the regular training period was started The criterion of learning was again set at 30 perfect consecutive trials In certain parts of the experiment the percentage of correct over incorrect responses was employed for reasons which will be indicated later.

The usual precautions were exercised in the control of all secondary cues. These included position, controlled through the experimental series, shadows, by the arrangement of the dim lights which illuminated the apparatus, sounds, through control tests conducted at various points in the experimental series; motivation, by regulating the time of experimentation, the amount of food given the animal, and the age of the subjects; handling of the animals (kinaesthetic and cutaneous), through control tests; and the various visual cues, through control tests to be described later.

In the first part of the experiment an attempt was made to find the point of subjective equality (*PSE*) for the red-blue stimuli. As has been pointed out in previous articles (4, 6, 7, 10), the brightness value of colored lights is quite different for the rat eye than for the human eye Consequently, when lights are equated for human vision they are not necessarily equated for the rat. Therefore, it is necessary either to run tests in which the brightness factor is varied in an irregular manner, or to let the animals determine for themselves the *PSE*. The latter procedure was done by lowering the brightness of one of the lights until a point was found at which the animals were only 50 per cent correct in their reactions. The second step was to take a new group of animals and to train them at the point of subjective equality Steps three and four were checks on the results obtained in steps one and two

RESULTS

Group A. Three animals having pigmented eyes were trained to discriminate red from blue, with red the positive color. The brightnesses of the two lights were equated for human vision by means of a flicker photometer. The results of this training are shown in Table 4. It is interesting to note that these correspond fairly well with the learning times obtained by Walton (9, p. 387). These animals were given from 100 to 120 extra trials in order that the brightness habit might be firmly established. Then a series of trials

TABLE 4
SHOWING RESULTS ON DETERMINATION OF THE *PSE*

Color pair (Red positive)	Rat (Sex)	Trials on learning[1]	Over-learning[2]	Trials to det *PSE*[3]
Red-blue	Norway (M)	105	120	650
Red-bue	Norway (F)	195	100	710
Red-blue	Cream (M)	161	120	725

[1] Criterion 30 perfect successive trials.
[2] Over-learning to insure establishment of brightness habit
[3] Blue light reduced in brightness.

was presented in which there was a gradual reduction in the brightness of the blue light. This control series was directed to the task of finding a point at which the rat could no longer discriminate on a brightness basis, the *PSE* There are some objections to regulating the brightness of the stimulus patches by rheostat methods. However, the validity of such objections is questionable as applied to the present experiment, for, if anything, the theory back of such criticism would favor the obtaining of negative results. In other words, according to this experiment the hue of the blue light should shift toward the red end of the spectrum, and thus, the *PSE* would be a point at which the animals could not distinguish between the two reds then being presented. This would make it impossible to obtain positive results in the control experiments. Hence, this criticism becomes invalid as applied to this particular part of the experiment. Figure 1 shows not only the learning curve for setting up the discrimination, but also the results of equating the lights for the animal's eye. It will be noted from the curve that the *PSE* was

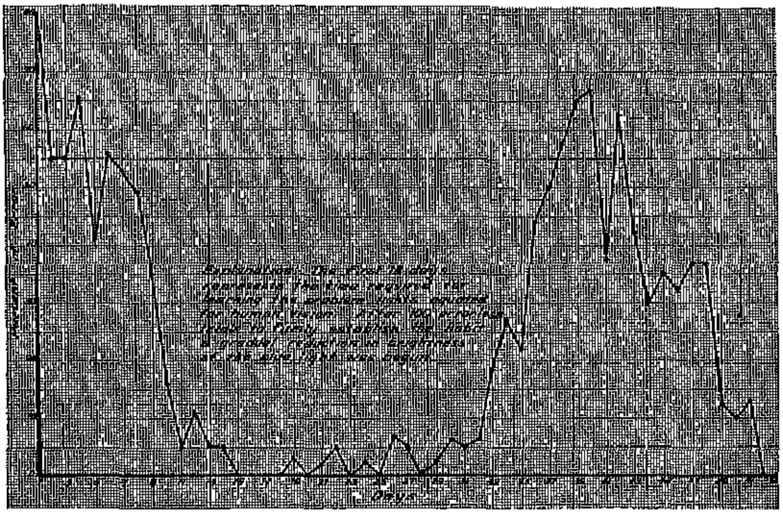


FIGURE 1
Showing Results Obtained in Determining the *PSE*, Red-Blue Discrimination

reached on the 20th day. To determine the exact candle power of the steps employed in this study, an extensive piece of work was done on the determination of the brightness values by means of the flicker photometer.[8] It is not necessary to describe in detail the standard photometric technique employed in this study. Table 5 shows the candle power of the red and blue lights at each step in brightness reduction.

TABLE 5

CANDLE POWER

(Meter-candles)

Step	Red light	Blue light
0	.04756075	.04756075
1	04206650	04647260
2	.03612231	.04372479
3	03148784	04084780
4	.02611321	03733230
5	.02134836	.03278500
6	.01847790	03043798
7	.01529510	.02326587
8	.01286766	.02156655
9	.00963800	01827451
10	00623270	01400283
11	.00528583	01143523
12	.00457623	.01021375
13	.00384082	.00709398
14	00320096	00577777
15	00215432	00480587
16	00147109	00386335
17	.00102812	.00321303
18	00081276	00231810
19	00053300	00225521
20	00036598	.00181468
21	00030404	.00139032
22	00021477	00084973
23	.00017236	00070960
24	.00009660	00051301
25	00003830	00036440
26	00000860	.00027744
27	00000200*	00021079
28	00000040*	.00008339
29		00005097
30		00003412
31		00000870
32		.00000392*

*Estimated.

---

[8] The writers wish to express their appreciation of the courtesy shown them by the department of physics at the University of Nebraska

The *PSE* value for red-blue is red = 0 and blue = 20 in terms of steps, or when red = .0475 meter-candles it is equal in brightness for the rat eye of blue at .0018 meter-candles

After the 40th day it will be noted that a reversal of the brightness habit took place,—the rats were then responding to the opposite color, i e, blue, as the less bright of the two Days 37 to 40 show a continued increase in errors, due to the establishment of this reversal

Group B. Since the rats which had been used in determining the PSE were trained to discriminate brightness differences and had had these habits well established early in the experiment, it was necessary to select a new group of animals for training at the *PSE*. One pair each of hooded and cream rats were selected for this experiment and trained at the *PSE* Table 6 shows the number of trials required for learning, as well as the additional trials given the animal after the criterion was satisfied It will be noted that the female cream rat is not included. This animal became so upset by the discrimination problem that it finally refused to run The same tendency had been exhibited by the male animal, hence the additional trials required for learning and, also, required for testing the establishment of the hue discrimination habit Figure 2 shows the learning time required for the three animals in terms of per cent of error Approximately 525 additional trials, as compared with the initial learning experiment, were required for learning at the *PSE*. When all of the various controls had been run, there was no doubt that these rats were distinguishing red and blue on the basis of hue

TABLE 6

SHOWING RESULTS OF TRAINING AT THE *PSE* RED-BLUE DISCRIMINATION

Rat (Sex)	Trials required for learning[1]	Additional trials	Per cent correct additional trials
Hooded (F)	652	48	100
Hooded (M)	648	52	100
Cream (M)	725	75	100

[1] Criterion 30 perfect successive trials

Group C As a check upon the possibility that some factor other than brightness might have determined the *PSE* in the first experiment, four animals were again trained on red-blue equated for human vision These animals mastered the problem in approximately 247

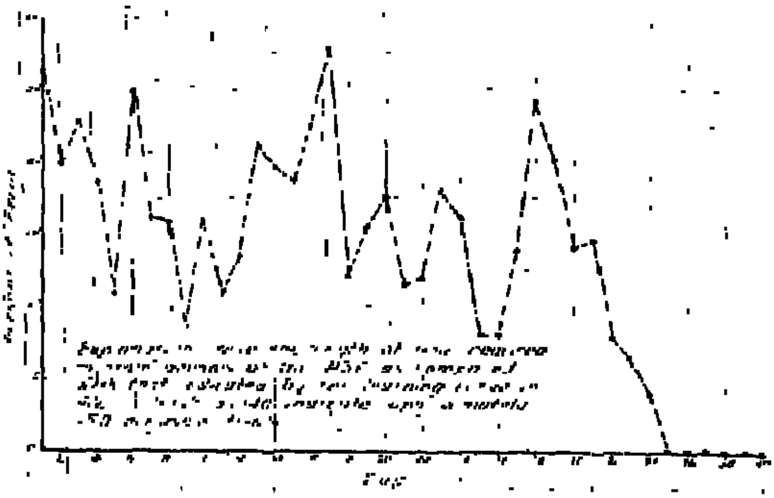


FIGURE 2
SHOWING LEARNING AND EVIDENCE OF RED-BLUE DISCRIMINATION, SUBJECTS TRAINED AT THE *PSE*

trials and were given the customary additional trials to establish the brightness habit. Then the brightness of red was reduced in the same manner in which the brightness of blue originally had been reduced. At no time during this reduction, covering approximately 900 trials, was there any point at which the rats failed to discriminate. Obviously then, the original *PSE* determinations were correct, at least so far as any secondary cue is concerned. In this series the weaker stimulus had been weakened, the values of the two lights still further separated.

Group D. There remained the possibility that the rats were either seeing only one color and that to them the pair was blue-darkness, or because one was monochromatic and the other multichromatic, some peculiar quality not visible to the human subject served in the place of hue. Both of these theories are not in accordance with the data shown in Figure 1. The reversal in reaction after the 37th day would be contrary to such views. However, since much has been made of the notion that the lower animals are either red-blind or at least red-weak, it seemed desirable to test some rats on red-darkness. For this study one hooded and three albino rats were selected. The results of the training and preliminary control experiments are shown

TABLE 7
SHOWING RESULTS OF RED-DARKNESS STUDY

Rat (Sex)	Trials required for learning[1]	Additional trials	Total trials reduction series[2]
Hooded (F)	114	36	725
White (F)	160	40	600
White (M)	213	37	625
White (F)	114	36	650

[1]Criterion 30 perfect successive trials
[2]Darkness positive stimulus

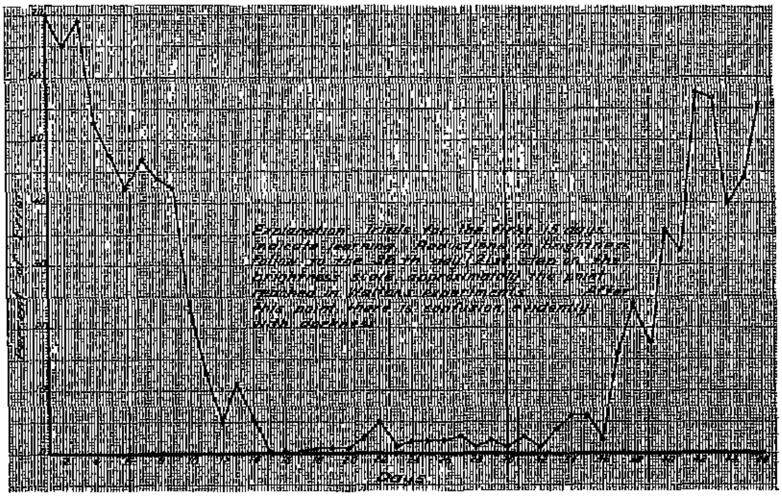


FIGURE 3
SHOWING EXPERIMENTS ON RED-DARKNESS DISCRIMINATION

in Table 7 Obviously, the rats distinguished the monochromatic red from darkness. After the preliminary training, occupying the first 15 days (Figure 3), the brightness of the red was lowered by gradual steps From the 15th to the 36th day the accuracy approximated 97 per cent. On the 36th day the 21st step on the brightness scale had been reached. This was identical with the point which the senior writer had reached in his first study, where it became necessary for him to place his hands in the runway in order that he might know which path had been selected Inasmuch as it had been impossible at that time to measure these brightnesses, this point was not stated in terms of meter-candles However, the study presented in Table 5 gives this point as .0003 meter-candles. Due to the fact that automatic doors had been installed, it was possible to go beyond this 21st step. Immediately, the errors began to increase, and on the 36th step there was complete failure As has been stated before, it was not the writers' intention to determine a limen. This study, however, indicates that it must lie somewhere between the 21st and 36th steps.

THE "WAVERING REACTION"

In 1930 Muenzinger and the senior writer[9] reported a reaction peculiar to discrimination problems After the preliminary learning period has passed (50 to 100 trials on a color discrimination problem with rodents as subjects), the animal appears to be comparing the two lights. To the observer it looks as though the overt movements are those which the animal *would* make *if* it were "judging" the lights. The animal simply sits at the partition and moves its body back and forth as if it were looking first at one light and then the other These movements appear as soon as the animal begins to attend to the lights as cues in the selection of the correct path and disappear as soon as the problem is mastered The writers' records indicate that when such movements are made during a trial, there are 88 chances in 100 that the response will be a correct one.

SUMMARY AND CONCLUSIONS

The Walton experiments were repeated in part, using a revised technique to eliminate the question of equivocal results Points of subjective equality were obtained for red-blue by lowering the brightness of the blue light. Then tests were run in which the brightness of the red light was lowered. New animals were then trained at the *PSE* and control tests run. Finally, a group was trained on red-darkness, and 0003 meter-candles was set as the point at which the response begins to weaken. The results lead to the conclusions that:

1. Rats can discriminate red-blue on the basis of wave length differences.

2. Rats can discriminate red from darkness and are rather sensitive, under semi-dark adaptation, to the red light used in the experiment.

---

[9] A meeting of the American Psychological Association at Iowa City, Iowa, December 28-30, 1930. See also (3, p 204); (9, p 391); (1, p. 178)

REFERENCES

1. COLEMAN, T B, & HAMILTON, W F. Color blindness in the rat *J. Comp Psychol.*, 1933, **16**, 177-181
2. LASHLEY, K S The mechanism of vision' V The structure and image-forming power of the rat's eye *J Comp Psychol*, 1932, **13**, 173-200
3. MUENZINGER, K F, & GENTRY, E Tone discrimination in white rats. *J Comp. Psychol*, 1931, **12**, 195-205
4. MUENZINGER, K. F., & REYNOLDS, H E. Color vision in white rats I Sensitivity to red *J Genet Psychol*, 1936, **48**, 58-71
5. MUNN, N L. An investigation of color vision in the hooded rat *J. Genet Psychol*, 1932, **40**, 351-362
6. ———. Further evidence concerning color blindness in rats *J Genet Psychol*, 1934, **45**, 285-302
7. MUNN, N. L, & COLLINS, M Discrimination of red by white rats. *J Genet Psychol*, **48**, 72-87
8. WALLS, G L. Visual cells of the white rat *J Comp Psychol*, 1934, **17-18**, 363-366
9. WALTON, W E Color vision and color preference in the albino rat II. The experiments and results *J Comp. Psychol.*, 1933, **16**, 373-394.
10. ——— Exploratory experiments in color discrimination *J Genet. Psychol*, 1936, **48**, 221-226

Department of Psychology
The University of Nebraska
Lincoln, Nebraska

The Journal of Genetic Psychology, 1938, 52, 183-201

A STUDY OF THE RELATIVE DIFFICULTY OF A PRIMARY READING VOCABULARY*

State Teachers College, Johnson City, Tennessee

LESTER R WHEELER

In selecting children's reading materials, educators are emphasizing more and more the importance of a vocabulary content appropriate for the ability levels of the children who are to use them Recent books for primary grades reflect the influence of the research of Gates, Thorndike, Horn and others in determining what constitutes an appropriate vocabulary Other investigators are attempting to determine the technicalities involved in efficiently teaching a basic reading vocabulary methods of presentation, adequate repetitions, phonetics, the value of word drill in or out of context, speed and accuracy of recognition, and the development of meanings are among some of the problems under study. Although there is considerable disagreement among teachers as to specific objectives in primary reading, all agree that the final test in reading depends upon recognizing word units in whatever sequence of phrase or sentence structure they may be found In the last analysis, good reading depends primarily on speed and accuracy of word recognition, and there is a close relationship between size of sight vocabulary and reading achievement.

In reviewing previous vocabulary studies, there appear two general types (*a*) lists of words compiled according to frequency of use, and (*b*) lists of words compiled according to difficulty. Most of the investigations fall under the first type, although varying considerably in methods of study A large number of the frequency lists have been compiled by analyzing reading materials and tabulating the number of times each word has occurred This technique first appeared as the basis for constructing spelling scales and around 1920 investigators began to use it to determine basic reading vocabularies. Among the early studies, were those of Housh (37) and Packer (46) who studied 10 first and 10 second grade readers. In 1924 Thorndike (63) published a list of 10,000 words found most

---

*Received in the Editorial Office on July 12, 1937

frequently in the general reading material for young people He used 45 different sources including textbooks, newspapers, correspondence, etc Ten years later he extended his list to 20,000 words, and it is still recognized as one of the most comprehensive studies in the field. In 1922 Selke and Selke (56) found that out of 1,636 words in 12 beginning books, only 38 words appeared in all 12 books Later these investigators analyzed 12 more primers and found similar results In 1925 Kircher (40) analyzed the vocabulary of 37 primers and first readers, and Horn, Horn and Packer (36), in cooperation with the National Society for the Study of Education, published a list of the most common words in the spoken vocabulary of pre-school children. This study is one of the first to determine which words the average first-grade child should be expected to know.

In 1926 Gates (23) published his *Reading Vocabulary for the Primary Grades,* revised in 1936 The 1,811 words in this widely used list are based on speaking and reading vocabularies, and most of the recent primary readers are checked against it. About the same time Ernest Horn (33) brought out a basic writing vocabulary of the 10,000 words most commonly used in writing outside the school. Four years later Beck (4) made a brief and incomplete analysis of first readers, and Wheeler and Howell (70) in 1930 compared the vocabularies of 10 primers and 10 first readers with Gates' list. They reported that while these readers show the centering effect of the Thorndike list, 26 per cent of the words did not appear in the first 500 of the Gates' list. Two years later the Child Study Committee of the International Kindergarten Union (35) published a study of the vocabulary of kindergarten children, based on the frequency of words used in conversation by children in the kindergarten, in the home, and while discussing pictures This list is widely used as a guide for constructing reading materials for beginners

In 1931 Harring (30) studied the vocabulary of 15 primers and found the percentage of overlapping small. In 1932 Faucett and Maki (21) compiled a list of words considered essential in the English language. This list, primarily a frequency list made from those of Thorndike and Horn, served as a basis, checked by Fitzgerald's list, for Durrell's *Vocabulary for Corrective Reading* (19). Fitzgerald (22), in 1934, examined children's letters and formulated a core vocabulary of words children use outside the school During the same year Gross (29) reported a study of the vocabulary of 10 pre-primers in which she found only four words *a, and, I,* and *the,* to be common to all 10 books In 1935 Stone (60) studied the new words appearing in 16 readers commonly used in the second grade, and pointed out the lack of standardization of vocabulary and the importance of skill in recognizing new words. Smith (58) investigated the papers which children wrote in various school subjects and showed the frequency of use, error, and grade in which the word was first used

During 1936 a committee including Faucett, Palmer, Thorndike and West (20) prepared a basic vocabulary for teaching English as a foreign language and classified the words according to the general experiences to which they apply Buchingham and Dolch (10) published a word list which combines their original free-association study with the lists of Gates, Thorndike, Horn, The International Kindergarten Union, and the spelling scales of Jones (39), Tidyman (62), Studley and Ware (61), Payne and Garrison (49), Bauer (2), and Horn (33). The frequency of use of 19,000 words were studied, and of this number over 10,000 were given grade placements according to usage Hockett and Neeley (32) made a comparison of the vocabularies of 33 recent primers. While the authors did not compile a word list, their findings are significant as to vocabulary burden in the widely used reading materials, showing the trend among recent publications to lighten the vocabulary load, but emphasizing the fact that the vocabulary range is still so great among primers that their use is materially limited as supplementary reading material.

Although various methods of studying the difficulty of reading materials have been discussed by Pressy and Lively (51), Patty and Painter (48), Lewerenz (43), Vogel and Washburne (68), Gray and Leary (25), Bear (3), Johnson (38) and others, each emphasizing more or less certain isolated factors in the total reading process, only a few investigators have attempted to compile a vocabulary ranked according to difficulty In 1928 Wiley (72) studied the rate of learning of two groups of children, equated by Binet *M.A.,* and placed under two teachers The first 60 words in the text were taught by each teacher in her own method of teaching word recognition Fifteen minutes a day was given to drill with flash-cards, and 20 minutes to reading twice a day The children were tested individually once a week for 12 weeks, and the words arranged according to difficulty on the basis of number of errors. The author concludes (*a*) the meaning a child has of a word greatly influences the difficulty of learning to recognize the word, (*b*) more attention should be given to the meaning in selecting vocabularies for first grade children, (*c*) children make use of cues in their attempt to recognize new words, (*d*) these cues are often inadequate and confusing, (*e*) inclusion of a word in the spoken vocabulary of children in children's literature, or in any of the present word lists does not seem to be an adequate measure of the learning difficulty of the word Wiley seems to be the only investigator who has attempted to control the factors of learning, methods of teaching, length of learning time, etc , in compiling a vocabulary list

In 1935 Rickard (53) reported a study to determine a basic recognition vocabulary of primary pupils, and ranked the words according to difficulty. He made a list of 119 words common to the first 200 in both M. D. Horn's and Gates' lists, and used it to test 207 children in Grades 1, 2, and 3 for visual recognition He studied the difficulty by ranking the words from easiest to hardest according to the number of pupils missing each, and discussed some of the factors affecting intrinsic difficulty. Rickard not only has a small number of cases per grade, but, when one studies the per cent of recognition by half grade in relation to the range, it is apparent that so large a percentage of the children in the second and third grades knew all the 119 words that a study of the relative difficulty would be determined largely by the few cases in Grade 1*A*. The technique he used to interpret his findings is questionable, his method of ranking faulty, and his method of testing from mimeograph sheets might materially affect the difficulty of recognizing the words.

Any attempt to determine the relative difficulty of a primary vocabulary presents many problems. Some of the conditions that produce difficulty have been studied by McGeoch and Oberschelp (44), Scott and Henniger (55), Thurstone (64), Cook (12), Woodrow (73, 74), Preston (52) and others. Their results indicate that intrinsic difficulty depends upon the interrelationship of various conditions such as configuration, letter span, spatial separation, illumination, and length of learning time Paterson and Tinker

(47), Gress (27, 28), Blackhurst (6, 7), Gates (24), and Dearborn (16) suggest that such typographical factors as style and size of type, leading, length of line, and margins may affect the difficulty of reading materials When one adds other variables such as mental and physical maturation, vision, audition, kinesthesia, emotions, sex differences, teaching techniques, language, and other factors, it becomes evident that as the conditions vary there will be changing levels of difficulty, and any generalized vocabulary list based upon the relative difficulty of the words may be materially affected by each one, or any varying combination of these factors Thus it is with certain reservations we attempt a study of the relative difficulty of a primary vocabulary, the ranking of any list is dependent upon so many variable factors that, at its best, it has many limitations even though we may assume that the large number of cases may partially take care of some of the variables.

Because the validity of a study of this type may depend to a great measure on controlling the factors of learning so as to eliminate as far as possible such varying conditions as vividness of presentation, frequency of repetition, typographical technicalities, length of learning time, type and quality of motivation, relation to context, degree of satisfaction, etc., we have used a new type of teaching material which, at least partially, controls these variables by incorporating the fundamentals of learning to read into a series of objectively controlled games for primary children.[1] The games, designed for teaching and drill, are constructed from Gates' *Vocabulary List for Primary Grades* *READ-O* series *IA* contains the 72 words most commonly used in children's reading material, series *IB* the 72 next most common, *IC* the next 72, etc , so that the series of three sets, or six games, contains 432 different words from the first 500 in Gates' list, arranged progressively throughout the series according to frequency of use As 43 of the 500 words are duplicated because of Gates' arrangement according to different parts of speech, the 432 different words cover Gates' first 475

The games are based on a flashcard technique. The teacher presents one flashcard at a time, calls the word, and uses it in a sentence from the context of the children's reading material. Each child plays the game on an individual card which has 16 of the 72 words arranged in four vertical and horizontal rows No two playing cards are alike and the 72 words are distributed equally over the 45 children's cards, thus controlling the frequency of occurrence As the teacher presents a flashcard and discusses the word to develop meanings, each child scans the 16 words on his card If he finds the word which has been presented, he puts a token over that word. The first child to get four tokens in a row calls out "*READ-O*" The teacher and pupils then check the child for accuracy by having him read aloud the four words in the row he has completed. If they check with the flashcards which are displayed in the holder in front of the class, the child wins the game. Through the game the children learn recognition and meaning of words, and the use of the words in phrase, clause, and sentence building, in a way which is closely integrated with the other reading material. Learning is highly motivated through the game and takes place incidental to a free-play activity. The efficiency of this method of teaching has been discussed by Wheeler.[2]

---

[1] *Read-O*, Series *IA* and *IB*, first half Grade I *Reado-O*, Series *IC* and *ID*, second half Grade I. *Reado-O*, Series *IIA* and *IIB*, Grade II Augsburg Publishing Co , Morristown, Tennessee.

READ-O, Series *IA,* consisting of the 72 words most frequently found in children's reading material was used in the present experiment. Data were gathered during 20 consecutive school days in October and November, 1936, from 227 children who had entered grade 1*B* the previous September. Six different public schools in various sections of Johnson City, Tennessee, and the Training School of the State Teachers College were selected in an attempt to get a fair cross-section of schools in this vicinity Seven teachers with four years' college training and several years' experience were chosen as representing the average public school teacher in our city systems. The classes were divided into experimental and control groups according to intelligence rating on the Dearborn *IA* Tests. Where a teacher had two sections, one was put into the control group and one into the experimental so as to equalize as much as possible any differences due to personalities and particular methods of the teachers. Also the experimental and control groups were equally divided between afternoon and morning sessions so that time of day would not influence the final results. No attempt was made to supervise or assist the teachers in any way other than instructing them as to the general procedure of playing the game with the experimental groups. In the experimental group *READ-O* was played 20 minutes a day during part of the regular reading period; otherwise both groups received the usual types of class-room instruction The repetition of the words and various other factors of the learning process were controlled during the game, although not controlled in the other school-room activities. The cooperation of all the teachers was excellent

---

[2] Wheeler, L R An Experimental Study of the Value of Informal Methods in Teaching Primary Reading *J Educ Res*, 1938, 31, 335-348. Data show that the experimental group of Grade *IB* children who played *READ-O* as a supplementary reading activity in the classroom made much greater improvement in vocabulary and general reading achievement than children of equal mental ability taught without using the games.

As there are no adequate standardized vocabulary tests available to test first grade so early in the school year, and as we needed an accurate check on the particular words used in the game, the best method of testing appeared to be oral recognition of words presented individually to the child The validity of this method of testing words out of context has been established by Monroe (45), and used by other investigators All the children in the experimental and control groups were tested in this manner at the beginning of the experiment, and at the end of 10 and 20 days The test consisted of the 72 words on the flashcards for *READ-O IA,* presented one at a time Interest was sustained by making a game of it, giving the child all the cards he could immediately recognize All the testing was done by students especially trained for this method, and the results tabulated according to the percentage of children immediately recognizing each word.

While the number of words analyzed in this study is small, they appear in all primary lists and thus we assume that they are basic in primary reading materials. The words are ranked in Tables 1 and 2 according to difficulty, on the basis of percentage of children who had learned each word according to Tests I, II and III. It is apparent that there is a close relation in rank between the words learned by the experimental and by the control groups, indicating that the method of teaching does not materially affect the ranking of the words, but does affect the rate of learning The experimental group learned more than twice as many words as the control group. The relative achievement of the two groups is indicated on Tables 1 and 2 by the percentile graphs. At the end of 20 days over half the words in the control, but only three in the experimental group,

TABLE 1

RANKING OF WORDS ACCORDING TO DIFFICULTY
Experimental group.

	Test I			Test II			Test III		Per-centiles
	%	Rank		%	Rank		%	Rank	
a	49	1.5	a	76	1.5	a	89	1	
dog	49	1.5	I	76	1.5	I	83	2	80
I	48	3	boy	67	3	girl	78	3	
boy	44	4	dog	63	4	dog	77	4	
to	41	5	girl	54	5	boy	76	5	70
girl	39	6	to	52	6.5	milk	68	6	
is	25	7.5	milk	52	6.5	run	67	7	
run	25	7.5	is	44	8.5	to	66	8	
play	20	9	bed	44	8.5	house	64	9	
house	16	10	run	39	10	father	62	10.5	
milk	15	11.5	father	38	11	is	62	10.5	60
mother	15	11.5	house	34	12.5	bed	59	12	
red	14	13	see	34	12.5	mother	58	13	
see	12	14	play	33	14	play	55	14	
father	11	15	big	32	15	see	54	15	
baby	10	16.5	mother	26	16.5	big	53	16	
like	10	16.5	red	26	16.5	red	52	17	50
the	9	18	baby	25	19	go	46	18	
water	8	19.5	in	25	19	baby	45	19	
bed	8	19.5	water	25	19	by	44	21.5	
eat	6	22	tree	23	21	it	44	21.5	
he	6	22	eat	22	22	tree	44	21.5	
home	6	22	look	21	23	up	44	21.5	
at	5	26.5	by	19	24.5	home	42	25	
big	5	26.5	up	19	24.5	in	42	25	
have	5	26.5	the	18	27	water	43	25	
no	5	26.5	home	18	27	the	41	27	
one	5	26.5	come	18	27	look	40	28	40
what	5	26.5	it	16	29	eat	38	29	
go	4	31	we	15	30	all	36	30	
has	4	31	he	13	32	like	35	31.5	
in	4	31	this	13	32	we	35	31.5	
am	3	37	of	13	32	come	34	33	
come	3	37	all	12	35	he	32	34	
do	3	37	go	12	35	am	30	36	
it	3	37	have	12	35	be	30	36	
on	3	37	are	11	37.5	saw	30	36	30
this	3	37	she	11	37.5	make	29	39	
up	3	37	us	10	39	of	29	39	
you	3	37	be	9	42.5	on	29	39	
tree	3	37	for	9	42.5	for	28	41	
an	2	44	make	9	42.5	my	26	42	
day	2	44	my	9	42.5	she	25	43	
for	2	44	on	9	42.5	no	22	44	
out	2	44	one	9	42.5	say	21	45	
of	2	44	an	8	47.5	are	20	47	
are	1	54.5	has	8	47.5	made	20	47	
be	1	54.5	me	8	47.5	one	20	47	20
by	1	54.5	you	8	47.5	an	19	50	
did	1	54.5	am	7	51.5	me	19	50	
get	1	54.5	made	7	51.5	what	19	50	
look	1	54.5	no	7	51.5	you	18	52	
may	1	54.5	what	7	51.5	has	17	54.5	
me	1	54.5	had	6	54.5	have	17	54.5	
my	1	54.5	saw	6	54.5	this	17	54.5	
not	1	54.5	say	5	56	us	17	54.5	
she	1	54.5	at	4	59.5	at	16	57	
saw	1	54.5	day	4	59.5	get	15	58.5	
say	1	54.5	did	4	59.5	was	15	58.5	
they	1	54.5	do	4	59.5	do	14	60	
us	1	54.5	not	4	59.5	from	13	63	
we	1	54.5	out	4	59.5	his	13	63	
all	0	66.5	from	3	64.5	had	13	63	
from	0	66.5	get	3	64.5	may	13	63	
her	0	66.5	like	3	64.5	not	13	63	
him	0	66.5	they	3	64.5	they	13	63	
his	0	66.5	was	2	67	did	12	67.5	
had	0	66.5	her	1	69.5	out	12	67.5	
made	0	66.5	him	1	69.5	day	11	69	10
was	0	66.5	his	1	69.5	him	9	70	
that	0	66.5	may	1	69.5	her	8	71	
make	0	66.5	that	0	72	that	7	72	0

TABLE 2
Ranking of Words According to Difficulty
Control group

	Test I %	Rank		Test II %	Rank		Test III %	Rank	Percentiles
									60
I	42	1	I	60	1	I	50	1	50
to	30	2	to	53	2	to	42	2	
a	25	3	a	49	3	a	41	3	40
boy	20	4.5	see	40	4	run	33	4	
see	20	4.5	run	35	5	boy	32	5	
girl	17	6	girl	31	6	see	31	6	
dog	16	7	boy	30	7	is	30	7	30
mother	15	8	dog	24	9	dog	26	8	
baby	13	9	is	24	9	look	25	9	
what	11	10	look	24	9	girl	23	10	
father	10	11.5	play	18	11	big	21	12	
home	10	11.5	father	16	13.5	house	21	12	
play	9	14	home	16	13.5	milk	21	12	
run	9	14	like	16	13.5	father	20	14	20
red	9	14	mother	16	13.5	in	19	16	
bed	7	18.5	baby	14	16	mother	19	16	
come	7	18.5	come	13	17	red	19	16	
one	7	18.5	milk	12	18	up	18	18	
the	7	18.5	red	11	19	come	17	19.5	
am	6	20.5	big	10	20	play	17	19.5	
is	6	20.5	am	8	21.5	baby	16	21.5	
house	5	23	house	8	21.5	like	16	21.5	
look	5	23	an	7	25	tree	15	23	
water	5	23	he	7	25	the	13	24	
all	3	26.5	it	7	25	it	12	25	
eat	3	26.5	the	7	25	what	11	26	
milk	3	26.5	what	7	25	go	10	28	
she	3	26.5	my	6	29	home	10	28	
an	2	37.5	they	6	29	we	10	28	10
are	2	37.5	you	6	29	for	9	30.5	
be	2	37.5	in	5	32	you	9	30.5	
for	2	37.5	saw	5	32	am	8	32.5	
get	2	37.5	tree	5	32	are	8	32.5	
go	2	37.5	are	4	35	have	7	35	
his	2	37.5	have	4	35	make	7	35	
he	2	37.5	not	4	35	no	7	35	
have	2	37.5	for	2	39	water	6	37	
it	2	37.5	make	2	39	bed	5	39.5	
like	2	37.5	no	2	39	he	5	39.5	
no	2	37.5	one	2	39	she	5	39.5	
not	2	37.5	she	2	39	saw	5	39.5	
of	2	37.5	at	1	49.5	by	4	44	
tree	2	37.5	all	1	49.5	my	4	44	
they	2	37.5	by	1	49.5	not	4	44	
up	2	37.5	day	1	49.5	on	4	44	
you	2	37.5	did	1	49.5	they	4	44	
by	1	53.5	do	1	49.5	all	3	49.5	
big	1	53.5	get	1	49.5	has	3	49.5	
did	1	53.5	her	1	49.5	may	3	49.5	
do	1	53.5	him	1	49.5	me	3	49.5	
her	1	53.5	me	1	49.5	of	3	49.5	
has	1	53.5	of	1	49.5	say	3	49.5	
in	1	53.5	say	1	49.5	at	2	56.5	
make	1	53.5	that	1	49.5	an	2	56.5	
may	1	53.5	up	1	49.5	did	2	56.5	
me	1	53.5	water	1	49.5	get	2	56.5	
my	1	53.5	was	1	49.5	one	2	56.5	
this	1	53.5	we	1	49.5	this	2	56.5	
was	1	53.5	bed	0	65.5	that	2	56.5	
we	1	53.5	be	0	65.5	was	2	56.5	
at	0	67.5	eat	0	65.5	be	1	62.5	
day	0	67.5	from	0	65.5	him	1	62.5	
from	0	67.5	go	0	65.5	his	1	62.5	
him	0	67.5	his	0	65.5	had	1	62.5	
had	0	67.5	had	0	65.5	day	0	68.5	
made	0	67.5	has	0	65.5	do	0	68.5	
on	0	67.5	made	0	65.5	eat	0	68.5	
out	0	67.5	may	0	65.5	from	0	68.5	
saw	0	67.5	on	0	65.5	her	0	68.5	
say	0	67.5	out	0	65.5	made	0	68.5	
that	0	67.5	this	0	65.5	out	0	68.5	
us	0	67.5	us	0	65.5	us	0	68.5	0

TABLE 3

PERCENTAGE OF RECOGNITION IN THE EXPERIMENTAL AND CONTROL GROUPS
Words by rank (for identification see Tables 1 and 2)

	Words by rank (*for identification* see Tables 1 and 2)					
		Experimental group			Control group	
	Test I	Test II	Test III	Test I	Test II	Test III
Grade	1B	1B	1B	1B	1B	1B
No of cases	158	142	149	79	76	75
No of words	72	72	72	72	72	72
%						
90-100	—	—	—	—	—	—
80- 90			1 to 3	—	—	—
70- 80	—	1 to 3	3 to 6	—	—	—
60- 70	—	3 to 5	6 to 12	—	1 to 2	—
50- 60	—	5 to 8 5	12 to 18	—	2 to 3	1 to 2
40- 50	1 to 6	8 5 to 10	18 to 29	1 to 2	3 to 5	2 to 4
30- 40	6 to 7 5	10 to 16 5	29 to 39	2 to 3	5 to 9	4 to 8
20- 30	7 5 to 10	16 5 to 24-5	39 to 50	3 to 6	9 to 11	8 to 16
10- 20	10 to 18	24 5 to 42 5	50 to 70	6 to 14	11 to 21 5	16 to 30 5
0- 10	18 through 72	42 5 through 72	70 through 72	14 through 67 5	21 5 through 65 5	30 5 through 68 5

TABLE 4

COMPARISONS OF EXPERIMENTAL GROUP, CONTROL GROUP, FREQUENCY AND DIFFICULTY RANKINGS

		No of words		r*
1	Comparisons within experimental group	72	Test I with Test II	78± 03
		72	Test I with Test III	79± 03
		72	Test II with Test III	.94± 01
2	Comparisons of experimental with control group	72	Test I (Exp) with Test I (Cont)	.78± 03
		72	Test II (Exp) with Test II (Cont.)	.66± 05
		72	Test III (Exp) with Test III (Cont)	87± 02
3	Comparisons with frequency lists	72	Test I with Gates' List	12± 08
		72	Test II with Gates' List	15± 08
		72	Test III with Gates' List	16± 08
		72	Test III with Thorndike's List	— 35± 07
		72	Test III with Wheeler & Howell's List	02±.08
		72	Test III with Fitzgerald's List	— 18±.08
		63	Test III with Gross' List	32± 08
		72	Test III with Horn's, E List	—29± 08
		72	Test I with Intern Kindergarten List	— 06± 08
		72	Test III with Intern Kindergarten List	— 08± 08
4	Comparing other difficulty lists with frequency	64	Rickard's with Horn's, M D.	39±.06
		64	Rickard's with Gates'	42± 07
		60	Wiley's with Gates'	21± 08
		60	Wiley's with Thorndike's	—.20± 08
5	Comparisons with difficulty lists	64	Test I with Rickard's	52± 06
		64	Test II with Rickard's	53± 06
		64	Test III with Rickard's	57± 06
		30	Test I with Wiley's	44±.10
		30	Test II with Wiley's	37± 11
		30	Test III with Wiley's	54± 09

*Rank difference method, Scott's formula.

are found in the first percentile Table 1 shows the small number of words recognized by the experimental group at the beginning of the experiment, and indicates a consistent increase in sight vocabulary as the experiment progressed. For example, at the beginning of the experiment half of the children, or 49 per cent recognized the word *a*, while 89 per cent recognized it at the end of 20 days. A study of the percentage of recognition and the ranking for the various tests gives an interesting picture of the relative difficulty of learning these 72 basic words Table 3 shows a similar picture, based on the percentile groupings of the children recognizing the various words.

A comparative study of the difficulty ranking and other word lists is shown by the correlations in Table 4. Part 1 shows a close and consistent relationship in rank between the different tests of the experimental group. While we are not able to say as to the relative position of the words which the control group failed to learn, a study of the ranks, quartiles, percentiles and correlations shows a striking similarity between the two groups in the relative difficulty of the words As all the correlations between the experimental and control groups are high, shown in Table 4, part 2, the lists supplement each other and strengthen our original ranking of the experimental group.

Many authors of primary readers have assumed there is a close relationship between frequency of use and difficulty of learning, and have used frequency lists as the basis for textbook construction with little or no regard to the intrinsic difficulty of learning the words. As a result teachers find primary readers do not give sufficient repetitions of the more difficult words to insure permanent learning, and the vocabulary burden is too heavy for the average child. The elementary curriculum is largely based on the theory of difficulty, i.e., as the child matures he is more and more capable of mastering increasingly difficult material, yet primary readers are not constructed on this basis. That there is practically no relationship between frequency of use and difficulty of learning is shown by the correlations in Table 4, part 3 Correlations between Tests I, II, and III and Gates' list indicate a negligible relationship. Gates' list, used widely as a basis for constructing primary reading material, does not indicate in any way the order of difficulty, and reading material based only upon frequency lists does not take care of the difficulty of learning The correlation between Thorndike's list and Test III gives a low negative correlation which indicates a slight reverse order between frequency and difficulty. There appears no relationship between Test III and Wheeler and Howell's frequency list of vocabularies of primers and first readers, and a negligible negative correlation between Test III and Fitzgerald's list which is compiled from children's letters written outside of school. A low positive correlation is found between Test III and Gross' frequency list of vocabularies of primers, and a low negative relationship appears between Test III and Horn's basic writing vocabulary, indicating a slight inverse relationship between difficulty of visual recognition and frequency of spelling usage. There is apparently no relation between difficulty of learning and the vocabulary of children entering the first grade as compiled by the International Kindergarten Union List. Summarizing the comparisons of our lists with these frequency lists, we find five positive and five negative correlations. The range of positive correlations is from .02 to .32 with an average of .12, and the range of negative correlations is from —.06 to —.35 with an average of —.19, indicating no definite positive relationship between our difficulty lists and these seven frequency lists. This is further substantiated by an analysis of the *PE*'s.

A comparison of other difficulty lists with frequency lists is given in Table 4, part 4. The correlation between Rickard's list, which is based on difficulty, and M. D. Horn's list, based on frequency, gives a low positive correlation. The correlation of Rickard's list with Gates' frequency list indicates a marked relationship, but this correlation loses some of its significance when we critically analyze Rickard's technique and method of study. The correlation of Wiley's list, based upon number of errors, with Gates' frequency list shows a negligible positive correlation; Wiley's list compared with Thorndike's shows a negative negligible correlation. A critical analysis of these four correlations fails to give evidence that there is a close relation between difficulty and frequency, and substantiates the previous comparisons.

Table 4, part 5, shows the correlations of Tests I, II and III with other studies that are based on difficulty of learning. The correlations of Tests I, II and III with Rickard's list is fairly high and increases from .52±.06 to .57±.06. Comparisons with Wiley's list also shows a positive relationship, the highest with Test III. It is unfortunate there is a small number of words common to both lists, and the large *PE* detracts from the significance of the correlation of Tests I and II with Wiley's list. It seems evident that, regardless of what constitutes the difficulty of learning, these three investigations have succeeded, in a way, in determining a difficulty rank of the words, even though approaching the problems by different methods, there is a closer correlation of our list with other difficulty lists than with the frequency lists studied. Also these correlations seem to substantiate our previous conclusion that in general the relative rank of the words is probably not greatly affected by methods of teaching. From a study of all the correlations it seems that studying relative difficulty through a controlled learning situation is probably a better technique than the methods previously employed

The value of a study of this type is largely determined by its practical applications It is however, only a sampling, or the beginning of the problem; we hope it may be extended to include a larger number of words and developed into a comprehensive, basic vocabulary ranked according to difficulty This study agrees with other investigators in indicating that word lists based on frequency of use are not ranked according to difficulty of learning. Such an erroneous assumption has caused writers to ignore the intrinsic difficulty of vocabularies in the construction of readers, tests, and reading materials for the lower grades. We believe that frequency and difficulty may be combined in the construction of reading materials Research on problem cases clearly indicates that vocabulary difficulty is one of the major factors causing reading problems. The difficulties of reading might be materially reduced by controlling the repetitions in the running words and increasing the number of new words according to the intrinsic difficulty of learning rather than relying only upon frequency of use This is an urgent and important task for research workers to consider. Teachers are fast recognizing the inefficiency of much of our present reading program, and are looking forward to a change in methods, or in reading materials, that will help them eliminate some of the present obvious difficulties.

SUMMARY

1 The majority of studies have been based on the frequency of use among children rather than on intrinsic difficulty of the words.

2. Many educators and textbook publishers have assumed that

the words most frequently used by children are the easiest to learn.

3 Experiments indicate that there are a number of factors which affect the difficulty of learning a primary vocabulary.

4 This investigation is the first which attempts to control the frequency of exposures of words and which teaches primary vocabulary through a directed free-play activity.

5 The results of this investigation show that the experimental group of children, taught through play, learn the vocabulary much faster than children taught by the regular school-room methods.

6 Although methods of teaching affect the rate of learning, they do not seem to materially affect the relative difficulty of learning the words, as indicated by the high correlations between the experimental and control groups.

7. There is also a high correlation between the relative ranking of the words at the beginning of the experiment and at the end of 10 and 20 days

8. There seems to be no significant positive correlation between our list based on difficulty of learning and lists compiled on the basis of frequency of use

9. The correlations of other studies, based on difficulty of learning, with frequency lists also indicate a low correlation.

10 There is a higher correlation between our list and other investigations of difficulty ranking than between our list and frequency lists, which further substantiates the conclusion that words most commonly used may or may not be the easiest for primary children to learn

11 There needs to be further research to compile more extensive difficulty rankings which may be used as a basis for the construction of children's reading materials.

REFERENCES

1. BAMBERGER, F E The effect of the physical make-up of a book upon children's selection *Johns Hopkins Univ Stud Educ*, 1922, No. 4. Pp. 162

2 BAUER, N The New Orleans public school spelling list New Orleans F. F Hansel & Bros, 1916

3. BEAR, M V Length of words as an index of difficulty in silent reading Master's thesis, Univ. Chicago, 1926. Pp. 77.

4. BECK, MAE M. The vocabulary of first readers *Peabody J Educ*, 1929, 6, 224-227

5. Betts, E. A. Reading disability correlates. *Education*, 1935, **56**, 18-25.

6. Blackhurst, J. H. Hygienic standards in type and format of reading materials. *Elem. Eng. Rev.*, 1928, **5**, 101-107.

7. ——— Size of type as related to readability in the first four grades. *School & Soc.*, 1922, **16**, 697-700.

8. Brown, E. M. *A master vocabulary in primary reading. 20th Yearbook, Dept. Elem. Sch. Prin., N. E. A.*, 1923, 296-306.

9. Buckingham, B. R. New data of the topography of textbooks. *30th Yearbook, Part II, Nat. Soc. Stud. Educ.*, 1931, 93-125.

10. Buckingham, B. R., & Dolch, E. W. A combined word list. Boston: Ginn, 1936. Pp. iv+186.

11. Carr, W. G. (Ed.). Better reading instruction. *Res. Bull. N. E. A.* 1935, **13**, 5.

12. Cook, T. W. The relation between amount of material and difficulty of problem solving. *J. Exper. Psychol.*, 1937, **20**, 178-185.

13. Cutright, P., Halvorson, G. P., & Brueckner, L. G. A study of one factor in the grade placement of reading materials. *Elem. Sch. J.*, 1928, **29**, 284-95.

14. Dale, E. Difficulties in vocabulary research. *Educ. Res. Bull.*, 1931, **10**, 119-122.

15. Dearborn, W. F. The psychology of reading. New York: Science Press, 1906.

16. ——— Difficulties in learning. Chicago: Univ. Chicago Press, 1932.

17. Dolch, E. W. A basic sight vocabulary. *Elem. Sch. J.*, 1936, **36**, 456-60.

18. ———. How much word knowledge do children bring to grade one? *Elem. Eng. Rev.*, 1936, **13**, 177-183.

19. Durrell, D. D. *A vocabulary for corrective reading. Elem. Eng. Rev.*, 1934, **11**, 106-110.

20. Faucett, Palmer, Thorndike, & West. Interim report on vocabulary selection for the teaching of English as a foreign language. London: King & Son, 1936. Pp. viii+506.

21. Faucett, L., & Maki, I. A study of English word values. New York: Oxford Univ. Press, 1932.

22. Fitzgerald, J. A. The vocabulary of children's letters written in life outside the school. *Elem. Sch. J.*, 1934, **34**, 358-371.

23. Gates, A. I. Reading vocabulary for the primary grades. Columbia Univ., Bureau Pub. Teach. Coll., 1926. Rev., 1936.

24. ——— What do we know about optimum length of lines in reading. *J. Educ. Res.*, 1931, **23**, 1-7.

25. Gray, W. S., & Leary, B. E. What makes a book readable? Chicago: Univ. Chicago Press, 1935. Pp. 352.

26. Gregory, C. A. The reading vocabularies of third-grade children. *J. Educ. Res.*, 1923, **7**, 127-131.

27. Gress, W. B. What shall the measure be? Is there an ideal length of line? *Amer. Printer*, 1929, **88**, 56-57.

28. ——— Legibility of type faces. *Amer. Printer*, 1929, **88**, 76-77.

29. Gross, A E A preprimer vocabulary study. *Elem Sch. J*, *1934*, **35**, 48-57

30 Harring, S. What primer shall I use next? *Elem. Sch. J*, 1931, **32**, *207-13*

31. Hill, M B A study of the process of word discrimination in individuals beginning to read. *J Educ Res.*, 1936, **29**, 473-482, 487-501.

32 Hockett, J. A, & Neeley, D. P A comparison of the vocabularies of thirty-three primers *Elem. Sch. J*, 1936, **37**, 190-203.

33. Horn, E A basic writing vocabulary. *Univ Iowa. Monog Educ.*, 1926, No 4.

34. Horn, M D. The thousand and three words most frequently used by kindergarten children *Childhood Educ*, 1927, **3**, 119.

35 Horn, M. D. (Chairman). A study of the vocabulary of children before entering the first grade. Child Study Com. Int. Kindergar. Un, 1928 Pp 36

36. Horn, E, Horn, M. D, & Packer, J L. The commonest words in the spoken vocabulary of the children up to and including six years of age *24th Yearbook, Part I, Nat. Soc. Stud Educ.*, 1925, 185-205.

37 Housh, E T Analysis of the vocabularies of ten second-year readers. *17th Yearbook, Nat Soc Stud Educ*, 1918

38. Johnson, G R An objective method of determining reading difficulty. *J. Educ. Res*, 1930, **21**, 283-287.

39. Jones, W F A concrete investigation of the material of English spelling. Univ. S Dakota, *1915*.

40. Kircher, H W Analysis of the vocabulary of thirty-seven primers and first readers *24th Yearbook Nat. Soc Stud Educ., Part I*, 1925.

41. Kyte, G C. Calibrating reading material *Elem Sch. J*, 1925, **25**, 533-46

42 Lebensohn, J. E The hygiene of reading *Ill Med. J.*, 1935, **68**, 425-29

43. Lewerenz, A S The measurement of the difficulty of reading materials. *Educ Res Bull*, 1929, **8**, 11-16.

44 McGeoch, J A., & Oberschelp, V J The influence of length of problem and of transfer upon rational learning and its retention. *J Gen. Psychol.*, 1930, **4**, 154-68

45 Monroe, M Methods for diagnosis and treatment of cases of reading disability. *Genet. Psychol Monog*, 1928, **4**, 333-456

46. Packer, J. L The vocabularies of ten first readers. *20th Yearbook, Nat Soc Stud Educ*, 1921

47 Paterson, D. G., & Tinker, M A Studies of typographical factors influencing speed of reading *J Appl. Psychol*, 1932, **16**, *605-613*; 1935, **19**, 647-651, 1936, **20**, 128-145.

48 Patty, W. W, & Painter, W I. A technique for measuring the vocabulary burden of textbooks. *J. Educ Res.*, *1931*, **24**, *127-134*.

49. Payne, B. R, & Garrison, S. C Payne-Garrison speller Chicago. Rand-McNally, 1931.

50 Peters, H N. The relationship between familiarity of words and their memory value *Amer. J Psychol*, 1936, **48**, 572-584.

51. Pressey, S. L., & Lively, B. A. "A method for measuring the vocabulary burden" of textbooks *Educ. Admin & Supervi*, 1923, **9**, 389-398

52 Preston, K. A The speed of word perception and its relation to reading ability. *J Gen Psychol*, 1935, **13**, 199-203

53 Rickard, G. E. The recognition vocabulary of primary pupils *J. Educ. Res*, 1935, **29**, 281-292

54. Rothlein, B. E The relative legibility of different faces of printing types *Amer J Psychol*, 1912, **23**, 1-36

55 Scott, T C., & Henniger, L L. The relation between length and difficulty in motor learning *J. Exper. Psychol.*, 1933, **16**, 657-678

56 Selke, E, & Selke, G. A A study of the vocabularies of beginning books in twelve reading methods *Elem. Sch J*, 1922, **22**, 745-749.

57. Shambaugh, C Q, & Shambaug, O L. A core vocabulary for elementary school pupils *J Educ Res*, 1929.

58. Smith, J. H The vocabulary of children based on an investigation of written papers selected from pupils in various subjects of the curriculum Oshkosh, Wis., State Teach. Coll., 1935. Pp 24.

59 Stanton, F. N., & Burtt, H. E The influence of surface and tint of paper on the speed of reading *J Appl. Psychol*, 1935, **14**, 683-93.

60. Stone, C. R. The second grade reading vocabulary *Elem. Sch J.*, 1935, **35**, 359-367.

61. Studley, C K, & Ware, A Common essentials in spelling. *Chico, Calif State Norm Sch Bull.*, 1914, No 7

62. Tidyman, W F. A survey of the writing vocabularies of public school children in Connecticut *U S Bur. Educ., Teach Leaflet*, 1921, No 15.

63 Thorndike, E L A teacher's word book of 20,000 words found most frequently and widely in general reading for children and young people Columbia Univ. Teach. Coll., 1931.

64. Thurstone, L L. The relation between learning time and length of task. *Psychol. Rev.*, 1930, **37**, 44-53

65 Tinker, M. A The effect of color on visual apprehension and perception. *Genet Psychol Monog*, 1932, **11**, 61-132.

66 ———. Illumination intensities for reading. *Amer. J. Ophthalmol*, 1935, **18**, 1036-1039

67 Trimble, O C., & Asperger, D H The "recognition spans" of "good" and "poor" readers: A comparative study. *J Appl. Psychol*, 1935, **19**, 665-682

68. Vogel, M, & Washburne, C. W. An objective method of determining grade placement of children's reading material *Elem. Sch. J.*, 1928, **28**, 373-381.

69. Ward, J L. Measuring vocabulary burden *Amer Sch. Board* J, **61**, 98.

70 Wheeler, H. E., & Howell, E. A A first grade vocabulary study. *Elem Sch. J*, 1930, **31**, 52-60.

71 Whipple, G. M. (Ed) The commonest words in the spoken vocabulary of children up to and including six years of age Report Nat. Com. Reading *24th Yearbook, Nat. Soc. Stud. Educ*, 1925, 186-198.

72 WILEY, W E Difficult words and the beginner *J Educ Res*, 1928, **17**, 278-289

73. WOODROW, H The measurement of difficulty. *Psychol Rev.*, 1936, **43**, 341-365.

74 ——— The interrelationship of conditions of difficulty. *J Gen Psychol.*, 1937, **16**, 83-131

75. WULTING, G Maturation as a factor in learning *Calif J Educ*, 1935 & 1936, **4**, 72-84, 148-164

State Teachers College
Johnson City, Tennessee

The Journal of Genetic Psychology, 1938, 52, 203-209

THE INFLUENCE OF MONTH OF BIRTH ON THE INTELLIGENCE TEST SCORES OF ADULTS*

Washington, D. C

H N. FIALKIN AND R. O. BECKMAN

Studies of the relationship of month of birth and intelligence have assumed two forms, depending upon the criterion accepted as the measure of intelligence. The most recent work has utilized eminence, as judged from *Who's Who in America, American Men of Science,* etc., as an indication of intelligence. The majority of the studies have, however, been based on intelligence test scores as direct measures of intelligence Comparisons of the data accumulated in this and other studies will be limited to work based on actual test scores

All work of the latter type which has hitherto been reported has been based on tests administered to children. The most extensive of these studies are those of Pintner in 1931, based on 4,925 test scores, and of Pintner and Forlano in 1933, based on 17,502 scores The latter definitely showed that month of birth is a factor in the intelligence test scores of children The present study is based on the test scores of 3,189 adults and makes possible a comparison of the results observed with children and those observed with adults.

The test scores utilized in this work were obtained from the case records of clients of the Adjustment Service of New York City, a free counseling service for adults which operated for 16 months. The tests used as diagnostic devices were the Pressey *Senior Classification and Verification Tests,* accepted as educational achievement tests and as indirect indicators of general intellectual capacity. The test scores were expressed in the records in terms of sigma ratings, on the basis of the established norms, and are presented as sigma ratings in this report. Since the norms for men and women differed, and since case records for men were far more common than those for women, the study was based on the ratings of men only The fact that a large number of the women failed to record the date of their birth was also a factor in the decision to limit the study to men.

---

*Received in the Editorial Office on July 13, 1937.

Statistics published by the Adjustment Service revealed that Pressey test ratings were available for 5,717 men. In order to obtain both the month of birth and the ratings of these men it was necessary to examine the individual case records. Since the records were filed by case number, irrespective of sex, the procedure adopted was to examine every case record, beginning with case number one and proceeding numerically, tabulating the test ratings only for men whose birth date was also available The tabulation was completed with 3,189 ratings which satisfied the above requirements. The method of selecting cases establishes month of birth as the only factor which could be of influence in the distributions of sigma ratings obtained for each month.

Comparing the 5,717 test ratings available and the 3,189 which were used in the study shows the smaller number to be a fair sample of the test ratings of all the men tested at the Adjustment Service

A word of elaboration concerning sigma ratings, also known as standard scores, is appropriate since the ratings used by the Adjustment Service differed somewhat from those commonly employed. Sigma ratings are raw scores which have been converted into comparative scores. They are expressed in terms of the standard deviation of the distribution of scores and represent the difference between specified raw scores and the mean accomplishment of the group examined The ratings usually employed vary from 0 (indicating an accomplishment equal to the average performance) in both a positive and negative direction. A scale ranging from 0 to 2.5 sigmas is generally sufficient to describe the accomplishment of the group tested, since a rating of plus 2.5 sigmas is above that of 99.4 per cent of the population, and that of minus 2.5 sigmas above that of only 0 6 per cent

The ratings used by the Adjustment Service differed from those commonly employed, in the fact that all ratings were expressed in positive terms. This was accomplished by setting 0 at minus 2.5 sigmas and 10 at plus 2.5 sigmas, the intervals between these extremes being expressed as the number of ½ sigma thus extended from minus 2.5 sigmas to minus 2 0 sigmas The mean of the general population is, in this terminology, at 5 0 ½ sigmas.

The skewed distribution of the ratings shown in Table 1 is due to the selected character of the persons who took advantage of the services offered by the Adjustment Service. A disproportionately

TABLE 1

omparative Distribution of Ratings of 5,717 Men and of Sample Used

	Total cases Number	%	Sample of 3,189 cases Number	%
1st ½ sigma	0	——	0	——
2nd ½ sigma	28	0 5	12	0 4
3rd ½ sigma	108	1 9	50	1 6
4th ½ sigma	362	6 3	197	6 2
5th ½ sigma	673	11 8	379	11 9
6th ½ sigma	917	16 0	526	16 5
7th ½ sigma	1070	18 7	620	19 5
8th ½ sigma	953	16 7	497	15 6
9th ½ sigma	1310	22 9	725	22 7
0th ½ sigma	296	5 2	183	5 7
	5717	100 0	3189	100 1
Mean score		6 60 ½ sigmas		6 61 ½ sigmas
St deviation		1 78 ½ sigmas		1 78 ½ sigmas

large percentage of the clients were from clerical and professional fields of work, and the group as a whole possessed more than the average amount of formal schooling.

On comparing the mean test scores of children born in the various months of the year, Pintner, 1931, and Pintner and Forlano, 1933, found that in general the means of the cold months tended to be lower than those of the warm months. Although it was impossible in the first of these studies to establish definitely reliable differences between the months of the various seasons, the increase in the number of cases involved in the second study made it possible to establish the difference between winter and all other seasons as reliable The difficulty encountered by all workers on the subject has been the fact that the absolute differences observed have been small The difference between the month with the highest mean and that with the lowest was, in both the above studies, less than 3 *IQ* points Large numbers of cases have been required to prove the reliability of the differences found.

The results obtained from the present study of the scores of adults in general confirm and agree with those obtained with children Working with a scale ranging from 1 to 10 only, instead of the extended *IQ* scale, the observed differences appear at first glance to be particularly small Although it has in most cases been impossible to prove that the observed differences are reliable, the fact that similar differences were found in three separate studies will be the basis of the argument that the differences are real.

The lowest mean rating was shown by adults born in February, the highest by those born in September. The absolute difference between these means is 0.37 ½ sigmas. The means and standard deviations for the individual months appear in Table 2.

Ranking the months in order of decreasing means and comparing with the ranking of the months shown by Pintner, 1931, and Pintner

TABLE 2

MEAN RATINGS OF ADULTS BY MONTH OF BIRTH

Month	Mean	SD	Number
January	6.61	1.80	275
February	6.39	1.69	277
March	6.59	1.78	281
April	6.64	1.84	240
May	6.73	1.76	255
June	6.69	1.79	274
July	6.67	1.82	292
August	6.56	1.76	262
September	6.76	1.63	256
October	6.60	1.75	263
November	6.50	1.71	244
December	6.62	1.79	270
Total	6.61	1.78	3,189

TABLE 3

RANKS OF MONTHS IN PINTNER'S, PINTNER AND FORLANO'S, AND THE PRESENT STUDY

Month	Pintner	Pintner-Forlano	Present study
January	11	11	7
February	10	12	12
March	9	9	9
April	6.5	3	5
May	5	6	2
June	3	1.5	3
July	4	5	4
August	8	8	10
September	2	1.5	1
October	1	7	8
November	12	4	11
December	6.5	10	6

Rho-Pintner, Pintner and Forlano	.535
Rho-Pintner, Present Study	.698
Rho-Pintner and Forlano, Present Study	.617

and Forlano, 1933, shows that in the latter study also, February *was the month with the lowest mean, while September shared the* highest ranking with June (Table 3).

It will be noticed that in general, throughout the three experiments, the warm and particularly the moderate temperature months rank higher than the cold months. The ranking of months in the study of adults shows a fairly strong relationship with the rankings obtained in the two studies with children The agreement in rank between the monthly means of adults and of children is, in fact, greater than that found between the means obtained in the two studies with children. The effect *of month* of birth *on intelligence* is reflected in the study of adults as well as of children

Blonsky, *in 1929*, reported that when he grouped the *IQ*'s of 453 tested children according to the time of the year in which they were born, the mean of those born in spring exceeded those for children born in other seasons, particularly those born in winter. Pintner in 1931 compared the means for the seasons of birth of his 4,925 cases and although he found that the mean for spring exceeded that for winter and autumn it was of the same magnitude as that for summer. The greatest difference, that between spring or summer and winter, did not, however, prove statistically significant. He *points out that Blonsky failed to give sufficient data to permit the* application of the test of significance to the differences he cited. The Pintner-Forlano *1933* study once again showed that the means of spring exceeded that of all other seasons Table 4 shows the means of the four seasons for the above cited experiments and for the present study. (Spring: April-June, Summer. July-September, Autumn· October-December, Winter: January-March)

TABLE 4

Means of the Seasons

	Blonsky	Pintner	Pintner-Forlano	Present study
Spring	84.3	97.20	102.35	6.69
Summer	81.5	97.20	102.06	6.66
Autumn	81.3	97.10	101.83	6.58
Winter	80.1	95.95	100.65	6.53

In each of the four studies the trend is, in general, the same. Although the difference between spring and winter is only 1.80 times its standard error in the present study, and only 1.74 times its standard error in the Pintner study, the fact that the mean of spring exceeds that of winter in each study lends significance to the observed differences. Moreover, the difference was shown to be definitely reliable in the 1933 study.

The importance of the fact that the results of several analogous experiments all tend in the same direction may be illustrated by R. A. Fisher's chi-squared method for obtaining a combined probability from individual tests of significance. Even if we limit the discussion to the two studies in which sufficient data were supplied but which failed to yield conclusive proof of the existence of a difference between spring and winter, an appreciable increase in the probability of such difference can be shown by this method. Calculating a combined probability from the inconclusive probabilities indicated by 1.80 and 1.74 times the standard error of the difference in the individual studies results in a probability which indicates that the true difference exceeds 2 times its standard error—considered a reliable difference by Fisher. Adding to this the fact that Blonsky's data were omitted from the above consideration and that the difference had been definitely shown to be reliable in the more extensive study lends assurance that the difference could in each study have been established as real if the number of cases had been sufficiently increased.

Seasons are rather unsatisfactory terms if we try to think of variations in intelligence in relation to variations in temperature. Months of moderate temperature occur both in spring and autumn, and autumn includes months of temperatures varying from moderate, as October, to markedly cold, as December. Selecting December, January, February and March as the coldest months of the year—based on mean temperatures in New York State—and April, May, September, and October as the most moderate months we can predict some degree of difference in favor of the moderate months from the fact that three of the four coldest months occurred in our previous classification *"Winter."* How will the

TABLE 5

MEANS OF MODERATE MONTHS COMPARED WITH THOSE OF COLD MONTHS

	Pintner	Pintner-Forlano	Present
Moderate	97.61	102.29	6.68
Cold	96.11	100.79	6.55
Ratio	2.37	4.06	1.72

difference between the means for cold and moderate months compare with that observed between spring and winter (Table 5)?

The absolute differences between moderate and cold months are in two of three studies slightly inferior to those found between winter and spring. The ratios of these differences to their standard errors are in general, however, of about the same magnitude as those shown for spring and winter. Considered individually, the differences observed in the Pintner study and the study of adults are not conclusive. By the logic previously postulated we can attach reliability to the difference observed in each of the experiments A difference then exists between the intelligence scores of persons born *in moderate months and those born in cold months*, as well as between those born in spring and those born in winter.

Conclusions

Month of birth has been shown to be a factor in the intelligence test scores of children. This study shows that this factor is also operative in influencing the test scores of adults.

Both children and adults who were born in spring score higher than those born in winter. In general, those born in months of moderate temperature are superior to those born in cold months.

Although month of birth has been established as a factor in intelligence test scores, it is a factor of but slight effect. The greatest *difference observed, that between the means of two individual months,* is less than 3 *IQ*, and in this study, in terms of sigma ratings, less than 0 5 ½ sigmas.

References

1. Bergen, G L, & Murphy, J. F Ten thousand clients of the adjustment service *Amer Asso Adult Educ.*

2. Fisher, R. A *Statistical methods for research workers* London Oliver & Boyd, 1925.

3. Pintner, R Intelligence and month of birth *J Appl Psychol.*, 1931, **15**, 149-154

4 Pintner, R., & Forlano, G. The influence of month of birth on intelligence quotients *J Educ Psychol*, 1933, **24**, 561-584

The Works Progress Administration
Washington, D C

The Journal of Genetic Psychology, 1938, 52, 211-225.

AN INVESTIGATION OF CONDITIONING IN CATS TO MULTIPLE STIMULI*[1]

Departments of Psychology, Columbia University and Brooklyn College

CHARLES NELSON WINSLOW, ROBERT KANTOR, CARL J. WARDEN

It may be open to question as to whether the phenomenon of conditioning the authors are about to report should be entitled "higher order conditioning" because of a point of departure from the usual higher order conditioning technique The unconditioned stimulus of food was employed as a means of direct reenforcement with conditioned stimuli beyond the first or "primary," which is contrary to the procedure described by Pavlov. Later in this paper the authors would like to bring up again for discussion the criteria of higher order conditioning and the delimitations of the concept. For the time being they will avoid the use of the term.

Pavlov and his associates (7) reported that they were unable to establish higher order conditioned responses beyond the third order. Finch and Culler (5) believe, however, that the failure of Foursikov (7), a co-worker of Pavlov, to obtain conditioned responses beyond the third order was probably due to lack of adequate motivation in the animal. After having reached the fifth order with shock applied to the paw of the dog as the unconditioned stimulus, they concluded: "There is every indication that we might proceed indefinitely to still higher orders. There is every reason to believe that higher order conditioning by this method is limited only by the number of available stimuli."

The method followed by the authors in establishing conditioning to multiple stimuli in cats was the motor alimentary conditioning method of Dworkin (2, 3, 4), in which the animal secures food upon raising the lid of a food box A similar method was used by Ten Cate (8) in conditioning cats to various auditory and visual stimuli.

---

*Received in the Editorial Office on July 22, 1937

[1] This investigation was carried out in the Laboratories of the Department of Neuropathology of the Psychiatric Institute of the State of New York, to the directors of which the authors wish to acknowledge their appreciation for the use of the facilities.

The unconditioned stimulus was food, and the response going to the food box and raising the lid to obtain the food

The apparatus consisted of a large wooden enclosure (Figure 1)

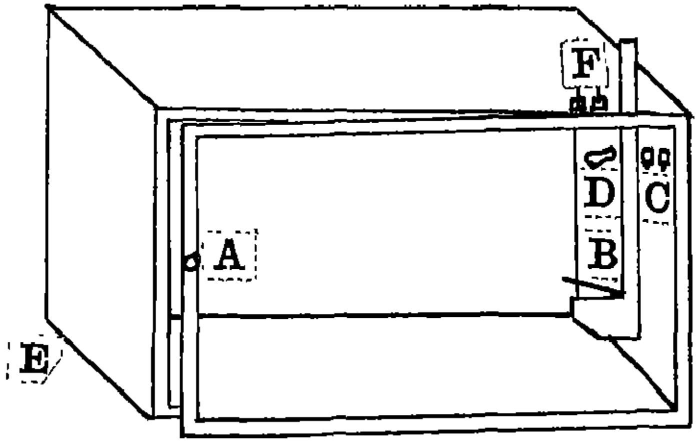


FIGURE 1

Diagram of the Conditioning Apparatus

Parts of the apparatus are labelled as follows: *A*—door of enclosure, *B*—food box and leader, *C*—bell and buzzer, *D*—red light, *E*—tubing for gust of air, *F*—switches for controlling visual and auditory stimuli

41½ x 31½ x 31 inches, insulated with one-inch fiber board to reduce sounds from the outside. The entrance door with its glass partition opened from one side. In front of the door was placed a one-way screen to shield the experimenter. A 25-watt bulb in the top of the enclosure supplied the illumination A monel metal leader connected the food box with the top of the enclosure and through this food was dropped by the experimenter after each correct response The lid of the food box had a protruding edge at the front which required first a downward and then an upward and forward movement of the cat's head. The devices for producing the five stimuli used in the course of the investigation were placed on the inside walls of the enclosure They were (*a*) electric bell, located at the right and just above the food box; (*b*) electric buzzer, side of the bell, (*c*) red light, at the left and just above the food box, (*d*) the light for the enclosure, which was extinguished to produce a conditioned stimulus; and (*e*) rubber tubing, which led to an aperture two inches above the floor in the end of the enclosure opposite the food box, so placed that it was possible to blow a gust of air onto the cat in almost every part of the enclosure. Electrical switches for control of the auditory and visual stimuli were conveniently placed outside on top of the enclosure. The apparatus was situated in a quiet and darkened room of the laboratory, about 40 feet from the living quarters.

The experimental situation provided two possible forms of incorrect behavior in the animal, in view of the fact that it was unrestrained in its movements within the enclosure. These were: (*a*) failure to go to the food box and raise the lid when the conditioned stimulus had been presented, and (*b*) going to the food box and raising the lid when the conditioned stimulus had not been presented One would say that the unconditioned stimulus was constantly present as a force, and in the hungry cat would produce a considerable amount of "undischarged tension," to use Lewin's expression, which would predispose it to go to the food box when the conditioned stimulus was absent. To limit the animal's going to the food box only when and if the conditioned stimulus has been presented is a more difficult situation than has usually been involved in conditioning work with animals.

The stimuli were presented in several different sequences with different animals in order to study possible effects of association of various pairs of stimuli in the series One animal, for example, was first given bell, second, light of the enclosure extinguished, third, blast of air, fourth, red light; fifth, buzzer Another animal received the five stimuli in the sequence bell, buzzer, light of enclosure extinguished, air, and red light

Ten trials were given at each experimental period daily The time interval between stimuli was varied from 5 to 30 seconds to prevent temporal conditioning. The interval between the conditioned stimulus and the unconditioned stimulus was 5 seconds. A first conditional stimulus was separated from a second conditioned stimulus by 6 seconds, when the association between two stimuli was being formed. The criterion for the establishment of conditioning to any one stimulus was 30 consecutive responses, viz , no failures to respond to the stimulus and no responses without the stimulus being presented. When the association between a new and an old conditioned stimulus was being formed to effect transition to the new, 10 consecutive correct responses to the new stimulus and none to the old served as the criterion for giving the new stimulus alone. The necessity for providing motivation for the animal as it was being conditioned to each new stimulus in such manner caused the experimenter to present food after the cat had responded to the new stimulus. This direct reenforcement was the departure from the usual procedure in higher order conditioning, referred to earlier in the paper.

The five cats employed as subjects ranged in age from six months to fifteen months. They were kept hungry for 24 hours preceding each series of trials, but during this time had access to water and milk. Pieces of salmon were used for the unconditioned stimulus, and were dropped by the experimenter into the pipe leading to the food box. To facilitate training at this point, the cover was removed at first, the cat simply learning to put its head into the food box to eat. After the animal did this with agility, the lid was restored, and 15 minutes a day for 7 to 14 days was spent in getting the cat accustomed to raising the lid. At first the experimenter by means of a string attached aided the animal in lifting the lid when it put its nose underneath to make the upward and forward movement. This proved to be a successful preliminary adaptation for all five animals, which were then ready for training with the conditioned stimuli.

The results for the five cats appear in Tables 1, 2, 3, 4, and 5, respectively. The second column in each table indicates the conditioned stimulus or the conditioned stimulus pairs employed in each two-day experimental period. The third column shows the number of responses to the conditioned stimulus as the numerator of a fraction, placed over the number of trials given as the denominator. Where stimulus pairs appear, the numerator indicates the number of responses to the first conditioned stimulus, i.e., the new stimulus that is being introduced during the process of transition. The fourth column shows the number of times within each experimental period that the animal went to raise the lid of the food box when no condititioned stimulus had been presented.

TABLE 1

Showing Results for Cat No. 1 with Each Conditioned Stimulus

Two-day testing period	Conditioned stimuli	No correct responses	No. responses without conditioned stimulus
1	bell	10/18	5
2	bell	12/20	4
3	bell	7/20	4
4	bell	18/20	2
5	bell	18/20	2
6	bell	19/20	2
7	bell	19/20	1
8	bell	20/20	0
9	bell	20/20	0
10	bell	17/17	0
11	buzzer and bell	0/20	0
12	buzzer and bell	11/20	0
13	buzzer	5/10	0
14	buzzer and bell	10/20	0
15	buzzer and bell	14/20	1
16	buzzer and bell	4/20	3
17	buzzer and bell	12/20	9
18	buzzer and bell	6/20	3
19	buzzer and bell	20/20	10
20	buzzer	20/20	12
21	buzzer	19/20	12
22	buzzer	20/20	11
23	buzzer	20/20	11
24	buzzer	20/20	5
25	buzzer and bell	10/10	10
26	buzzer	18/20	3
27	buzzer	20/20	0
28	light and buzzer	6/20	5
29	light and buzzer	16/20	7
30	light	20/20	7
31	light and buzzer	7/20	0
32	light and buzzer	7/10	3
33	light	19/20	3

These results for individual animals are summarized in Table 6, which shows the number of trials required for conditioning to each stimulus alone in the series. Table 7 supplements Table 6 by showing the number of trials required to reach the norm of 10 consecutive correct responses to the new stimulus when the new and the old were associated. The averages from Tables 6 and 7 combined to-

TABLE 2

SHOWING RESULTS WITH CAT NO. 2 WITH EACH CONDITIONED STIMULUS

Two-day testing period	Conditioned stimuli	No. correct responses	No responses without conditioned stimulus
1	bell	16/20	6
2	bell	20/20	2
3	bell	20/20	0
4	bell	10/10	0
5	buzzer and bell	3/20	0
6	buzzer and bell	18/20	0
7	buzzer and bell	10/10	0
8	buzzer	9/10	0
9	buzzer and bell	14/20	0
10	buzzer	4/10	0
11	buzzer and bell	9/20	6
12	buzzer and bell	11/20	3
13	buzzer and bell	3/20	1
14	buzzer and bell	4/20	0
15	buzzer and bell	6/20	0
16	buzzer and bell	4/20	0
17	buzzer and bell	3/20	0
18	buzzer and bell	3/20	0
19	buzzer and bell	5/20	0
20	buzzer and bell	8/20	4
21	buzzer and bell	4/20	0
22	buzzer and bell	2/20	0
23	buzzer and bell	18/20	0
24	buzzer	20/20	2
25	buzzer	10/10	0
26	light and buzzer	8/20	0
27	light and buzzer	17/20	0
28	light	19/20	3
29	light	20/20	4
30	light	20/20	1
31	light	20/20	0
32	air and light	8/20	11
33	air and light	10/20	2
34	air and light	9/20	12
35	air and light	11/20	2
36	air and light	18/20	0
37	air and light	10/10	0
38	air	20/20	1
39	air	20/20	0
40	air	20/20	0
41	air	20/20	0
42	red light and air	10/20	3
43	red light and air	18/20	2
44	red light and air	10/10	0
45	red light	15/20	0
46	red light	19/20	1
47	red light	20/20	0

TABLE 3

Showing Results for Cat No. 3 with Each Conditioned Stimulus

Two-day testing period	Conditioned stimuli	No. correct responses	No. responses without conditioned stimulus
1	bell	10/16	5
2	bell	17/20	4
3	bell	19/20	2
4	bell	17/20	1
5	bell	18/20	1
6	bell	8/10	0
7	bell	20/20	0
8	buzzer and bell	5/20	1
9	buzzer and bell	10/20	4
10	buzzer and bell	6/15	0
11	buzzer and bell	10/10	0
12	buzzer	16/20	0
13	buzzer and bell	4/20	5
14	buzzer and bell	10/20	3
15	buzzer and bell	4/20	5
16	buzzer and bell	5/20	2
17	buzzer and bell	10/10	0
18	buzzer	11/20	0
19	buzzer and bell	10/10	0
20	buzzer	17/20	1
21	buzzer	8/10	0
22	buzzer and bell	10/17	3
23	buzzer and bell	17/20	3
24	buzzer and bell	8/20	2
25	buzzer and bell	10/10	0
26	buzzer	18/20	1
27	buzzer	20/20	1
28	buzzer	10/10	0
29	light and buzzer	10/20	5
30	light and buzzer	18/20	0
31	light	20/20	3
32	light	20/20	2
33	light	20/20	2
34	light	20/20	0
35	light	20/20	0
36	air and light	7/20	5
37	air and light	8/15	2
38	air and light	9/20	0
39	air and light	4/17	0
40	air and light	7/20	0
41	air and light	16/20	0
42	air	16/20	0
43	air	20/20	2
44	air	20/20	0
45	red light and air	12/20	2
46	red light and air	14/20	1
47	red light and air	10/10	0
48	red light	20/20	0
49	red light	16/16	0
50	red light	20/20	0

TABLE 4

SHOWING RESULTS FOR CAT NO. 4 WITH EACH CONDITIONED STIMULUS

Two-day testing period	Conditioned stimuli	No. correct responses	No. responses without conditioned stimulus
1	bell	12/20	10
2	bell	14/20	7
3	bell	15/20	5
4	bell	14/20	4
5	bell	9/20	6
6	bell	15/20	2
7	bell	12/20	2
8	bell	12/20	2
9	bell	7/20	4
10	bell	6/20	2
11	bell	12/20	4
12	bell	7/10	5
	30 day interval during pregnancy		
13	bell	9/20	1
13	bell	13/20	2
15	bell	14/20	2
16	bell	17/20	5
17	bell	18/20	0
18	bell	11/15	1
19	bell	15/20	0
20	bell	13/20	0
21	bell	11/15	3
22	bell	15/20	7
23	bell	20/20	2
24	bell	20/20	0
25	light and bell	10/20	3
26	light and bell	10/10	0
27	light	20/20	0
28	light	20/20	0
29	light	10/10	0
30	air and light	15/20	2
31	air	20/20	2
32	air	20/20	0
33	air	10/10	0
34	red light and air	17/20	3
35	red light and air	19/20	1
36	red light	20/20	2
37	red light	20/20	0
38	buzzer and red light	7/15	0
39	buzzer and red light	18/20	0
40	buzzer and red light	18/20	0
41	buzzer	20/20	0
42	buzzer	20/20	0
43	buzzer	10/10	0

TABLE 5

Showing the Results for Cat No 5 with Each Conditioned Stimulus

Two-day testing period	Conditioned stimuli	No correct responses	No responses without conditioned stimulus
1	buzzer	5/12	3
2	buzzer	8/20	3
3	buzzer	7/20	3
4	buzzer	1/13	0
5	buzzer	12/20	2
6	buzzer	13/20	1
7	buzzer	14/20	3
8	buzzer	14/20	6
9	buzzer	13/20	4
10	buzzer	12/20	0
11	buzzer	7/20	1
12	buzzer	11/20	1
13	buzzer	10/20	0
14	buzzer	12/20	0
15	buzzer	10/17	0
16	buzzer	17/20	1
17	buzzer	10/15	1
18	buzzer	20/20	3
19	buzzer	20/20	0
20	buzzer	10/10	0
21	light and buzzer	8/16	1
22	light and buzzer	10/20	0
23	light and buzzer	9/20	4
24	light and buzzer	10/20	0
25	light	20/20	0
26	light	20/20	0
27	air and light	11/20	0
28	air and light	15/20	4
29	air and light	10/10	0
30	air	20/20	8
31	air	20/20	4
32	air	20/20	0
33	bell and air	14/20	6
34	bell and air	15/20	0
35	bell	20/20	2
36	bell	20/20	0
37	red light and bell	12/20	0
38	red light and bell	18/20	1
39	red light	20/20	2
40	red light	20/20	0
41	red light	10/10	0

TABLE 6

Showing the Number of Trials Required by Each Cat for Conditioning to Each Stimulus Presented Alone in the Series

	Stimuli				
Cat	I	II	III	IV	V
1	195	150	40		
2	70	50	80	80	60
3	126	120	100	60	56
4	460	50	50	40	50
5	367	40	60	40	50
Average	244	82	66	55	54

TABLE 7

Showing the Number of Trials Required by Each Cat to Effect Transition from an Old to a New Stimulus when Stimuli Are Paired

	Stimulus pairs			
Cat	I and II	II and III	III and IV	IV and V
1	160	70		
2	320	40	110	50
3	227	40	112	50
4	30	20	40	55
5	66	50	40	40
Average	161	44	75	49

TABLE 8

Showing the Total Average Number of Trials Required for Conditioning to Each Stimulus

Stimulus	I	II	III	IV	V
No trials	244	243	110	130	103

gether in Table 8 give the total number of trials required to obtain conditioning to each of the five stimuli. Cat No. 1 died after it had been conditioned to three of the five stimuli.

The results in general show that conditioning the animal to the first or "primary" stimulus prepared it for conditioning to all the other stimuli which followed, as is demonstrated by the comparatively few trials needed to establish conditioning to each new stimulus. This finding would substantiate the theory that conditioning is a general modification of the organism as a whole, and not the modification of a particular set of neurones.

That the number of trials required in passing from the older conditioned stimulus to the new conditioned stimulus is in part a function of what those stimuli are is revealed in the results Cats Nos. 1, 2, and 3 which were given Bell and Buzzer as the first two stimuli required an average of 236 trials to effect the transition from the first stimulus to the second in the same sense modality, compared with an average of 48 trials for cats Nos 4 and 5 which received the sound of the buzzer as the first stimulus, and the light of the enclosure extinguished as the second, two stimuli from different sense modalities. In the later transitions where a disproportionately large number of trials was required, temporal conditioning was the cause, and had to be broken down before the experimenter could continue with the conditioned stimuli.

In the earliest stage of conditioning, going to the food box and raising the lid when the conditioned stimulus had not been given occurred frequently. This form of incorrect response disappeared later than the other type, that of failure to respond when the conditioned stimulus had been given There was a marked tendency for the former type to appear when the cat was receiving the pairs of associated stimuli. Signs of temporal conditioning were particularly apparent at these times, and were not present until conditioning to the primary stimulus was almost established. This is further support for the authors' contention that conditioning the animal to the primary stimulus prepared it for more rapid conditioning to any subsequent stimuli, time being one of them

The behavior of the animals between trials underwent a marked change with the progress of conditioning. In the early stages the cats would frequently wash, scratch at the walls of the enclosure, and show considerable lack of orientation to the food box. When the conditioned stimulus was presented, the response, although occurring quite soon, did not occur with the spontaneity that characterized it in the later stages of conditioning By the time conditioning was established to the first stimulus each animal was showing an objective orientation to the goal between trials Many times it would stand almost motionless at the end of the enclosure opposite the food box, gaze fixedly at the box, and then dash forward as soon as the stimulus had been presented. Dworkin reports that this latent period between the stimulus and the response of the cat dropped from 15 seconds to 2 to 5 seconds as conditioning progressed It frequently happened that an animal would make a complete circuit of the box following each response, the movement ending with the animal facing the box, set for the next response.

Certain qualitative differences in the individual animals seem to correlate rather closely with the quantitative differences reported. Cat No. 5 could be described as highly tense and vivacious It was very active during the tests, rapidly moving around the enclosure between trials, with its tail erect and body alert It was more rapidly conditioned than the more phlegmatic ones. The pregnancy of cat No. 4, necessitated an interruption of 30 days during the process of conditioning her to the first stimulus. It will be observed from Table 4 that after this interval training was resumed with practically no loss in performance as a result of the inactivity.

In view of the fact that the unconditioned stimulus of food was presented with each of the multiple stimuli as a means of direct reenforcement, the authors have hesitated to use the term "higher order conditioning" at all. There can be little doubt that Pavlov maintained that the unconditioned stimulus should not be presented at all after its use in the establishment of the primary or first order conditioned response. He says in this connection:

> With the help of this strong conditioned stimulus (beat of metronome) it has been found possible to give another stimulus with conditioned properties like the first For if some new and more or less neutral stimulus is applied with the metronome alone, i e., *not at the same time giving food,* this new stimulus also acquires the character of an alimentary conditioned stimulus. Conditioned reflexes established in this manner are termed "secondary conditioned reflexes" The essential condition is *that the new stimulus should be withdrawn some seconds* before the primary stimulus is applied.

It would seem that failure of Foursikov to establish higher order conditioned reflexes beyond the third order was due to the operation of extinction of the primary reflex by the time the third order was reached In other experiments Pavlov showed that extinction proceeded rapidly, so that it would be difficult to introduce many higher order stimuli before the primary conditioned reflex became extinct. Finch and Culler (6) seem to have been aware of the need for reenforcement with the unconditioned stimulus in establishing conditioned responses beyond the primary, in order to maintain motivation in the animal They argue that they were giving the reenforcement indirectly, however, by applying the electric shock to the thorax of the dog, after having first applied it to the forepaw to obtain the primary conditioned response of withdrawal of the leg. They believe that with the use of such reenforcement they were able to preserve a constant motivation in the animal Admittedly, the authors of the present paper gave the reenforcement directly, and in that way precluded the possibility of extinction of the conditioned response while attempting to establish it to other stimuli beyond the first.

Although the authors would not argue that the conception or even the criteria of higher order conditioning should be changed, they would like to emphasize the limitation in higher order conditioning that without reenforcement in some way with the unconditioned stimulus, the conditioned response tends rapidly to become extinct.

The authors would conclude for cats, as have Finch and Culler from their work with dogs, that conditioning can be established to many stimuli beyond the first If, however, the phenomenon of conditioning to multiple stimuli which has just been described cannot appropriately be called "higher order conditioning," there is the question as to what term should be employed for it

SUMMARY AND CONCLUSION

The experiment consisted of conditioning cats to multiple stimuli by the Dworkin motor alimentary conditioning method, in which food is obtained by the animal when it lifts the lid of a box after a conditioned stimulus has been presented The stimuli used were bell, buzzer, red light, source of illumination of the enclosure extinguished, and a gust of air blown onto the animal Errors were of two types in this experimental situation, namely, failure to go to the food box and raise the lid when the conditioned stimulus had been given, and going to the food box when the stimulus had not been given The authors found that the latter type was more difficult to eliminate than the former.

The results indicated that conditioning was established more and more rapidly to each stimulus subsequent to the primary stimulus, suggesting that the conditioning process involves a general modification of the animal. Furthermore, conditioning was more rapid when two consecutive stimuli belonged to different sense modalities than when they were from the same. Variations of the sequence of stimuli with different animals made it possible to determine this.

The authors raise the question as to whether the phenomenon of conditioning to multiple stimuli which they obtained should be labelled "higher order conditioning," in view of the fact that they directly reenforced the response to each consecutive conditioned stimulus with the original unconditioned stimulus of food, contrary to the procedure for higher order conditioning outlined by Pavlov They point out, however, that Finch and Culler (6) indirectly reenforced the response to conditioned stimuli beyond the primary by applying the electric shock to a different part of the dog's body from that used to establish the primary response, and called the phenomenon "higher order conditioning" The writers conclude with the question as to whether the use of the term can be extended to include conditioning to multiple stimuli when the conditioned stimuli beyond the primary are directly reenforced by the unconditioned stimulus, or whether another term should be chosen for this phenomenon.

REFERENCES

1. BROGDEN, W J., GIRDEN, E, METTLER, F. A., & CULLER, E Acoustic value of the several components of the auditory system in cats. *Amer. J. Physiol*, 1936, **116**, 252-261
2. DWORKIN, S. Conditioned motor reflexes in cats. *Amer. J. Physiol.*, 1934, **109**, 31
3. DWORKIN, S, SEYMOUR, S I, & SUTHERLAND, G F. The conditioned response method of testing hearing in cats. *Quart. J Exper Physiol*, 1934, **34**, 23-30
4. DWORKIN, S. Alimentary motor conditioning and pitch discrimination in dogs *Amer J Physiol.*, 1935, **112**, 323-328
5. ———. Pitch and intensity discrimination by cats *Amer J Physiol*, 1935, **112**, 1-4.
6. FINCH, G, & CULLER, E. Higher order conditioning with constant motivation *Amer. J Psychol*, 1934, **46**, 596-602.

7 PAVLOV, I P. Conditioned reflexes (Trans Anrep) Oxford: Oxford Univ. Press, 1927. Pp 430

8. TEN CATE, J Akustische und optische Reaktionen der Katzen nach Teilweisen und Totalen Exstirpationen des Neopalliums *Arch. neerl de physiol*, 1934, **19**, 191-264.

Brooklyn College
Brooklyn, New York

Columbia University
New York City

Columbia University
New York City

SHORT ARTICLES AND NOTES

The Journal of Genetic Psychology, 1938, 52, 227-231

A JUMPING METHOD FOR ESTABLISHING DIFFERENTIAL RESPONSES IN PIGEONS*[1]

Otho S A Sprague Memorial Institute and Division of Psychiatry, University of Chicago, and Northwestern University

WARD HALSTEAD[2] AND GEORGE YACORZYNSKI

That the establishment of differential responses in animals depends very often on the use of an adequate training technique has been well illustrated by Lashley's (5) work. Prior to the development of the jumping method in training rats, it was generally assumed that rodents could not form pattern discriminations In pigeons, Gundlach (1) was able to obtain differential responses by the use of a modified Watson-Yerkes discrimination box only after laborious training He states. " . there was such difficulty in setting the problem such that these seemingly stupid creatures could reach any adequate criterion of learning." Hamilton and Coleman (2), and Hamilton and Goldstein (3), on the other hand, attacking an analogous problem, but using a jumping method similar to that employed with rats, were able to obtain differential responses in pigeons in 10 to 100 trials Their pigeons were trained to jump from a platform to one of two perches, one perch facing the negative, the other the positive stimulus The perches were located before a wall in such way as to provide inadequate space for the animal actually to perch Thus jumping to the wrong perch was punished since the animal soon lost its balance and fell into a net However, jumping onto the correct perch automatically opened a door in the wall, permitting the animal to enter a dark compartment The authors do not mention what method was used to get the animals

---

*Received in the Editorial Office on April 8, 1937

[1] The paper is a preliminary report of one of a series of investigations conducted in the laboratory of Dr Franklin Fearing concerning the behavior changes following lesions in the central nervous system.

[2] Fellow of the National Research Council

to jump in the first instance, but simply indicate that prior to the experiment proper, the animals had been trained to jump

In making comparative studies of this nature, the factor of motivation is of prime importance. For avian subjects hunger motivation is not satisfactory (in part because of the nature of the digestive apparatus, consisting of a crop in which the food is stored and a stomach proper where digestion is started), although Layman (6) and others report varying degrees of success after training running into hundreds of trials. While Coleman and Hamilton successfully utilized a darkness preference as an incentive in normal pigeons, in studies involving lesions within the central nervous system, this procedure is probably limited in application since there is reason to believe that one effect of such lesions may be a reversal of preferences (4)

For establishing differential responses to brightness and pattern in pigeons, we have developed a jumping method which represents a modification of that employed by Hamilton and Coleman, but which yields a quantitative index of obstruction or motivation associated with the jumping of the animal. The apparatus consists essentially of a variable speed, mechanically rotated perch upon which the bird is placed, and two platforms upon which the bird may jump in leaving the moving perch The perch, made of round, soft wood, ⅞ of an inch in diameter and 16 inches long, is elevated 43.5 inches from the floor. The ends of the perch are closed with galvanized iron The perch is belt driven from a 30 to 1 reduction gear by means of a ¼ *HP DC* motor, controlled by a bank of parallel-coupled resistances. Through appropriate switches the perch may be accelerated by discrete steps from rest to the maximum speed, necessary for our purposes, of 18 *RPM*. At this highest speed no bird was able to stay on the perch and either jumped to one of the platforms or fell to the floor

Two and a half inches in front and two inches below the perch are two platforms of equal size (7.5" x 5") separated by a distance of approximately three inches. The end of each platform nearest the perch is hinged. By means of a trigger arrangement both platforms may be set either to support the weight of the animal or to trip, causing the animal to fall to the floor[8]. Our stimulus objects are black or white background cards, 7.5" x 5" in size, on each of which is one of three figures: a square, a triangle and a circle, all with an area of four square inches. Two identical cards are used on each platform, one horizontally and one vertically located at the distal end With the stimulus cards in the horizontal position only it was observed that differential responses did not appear.

---

[8] In a more recent arrangement, the bird falls into a chute which in turn delivers it into a net.

The apparatus is inclosed by dark cloth to provide a homogenous environment and thus to avoid utilization of incidental factors by the animal The apparatus is illuminated by means of a 60 watt frosted lamp placed 18 inches directly above the platforms

Our preliminary training procedure was as follows: the wings of the bird were first taped behind the back to prevent flight, the bird was then placed on the perch with its head directed towards the two platforms, both of which were set to support the weight of the bird; the perch was then started at its slowest speed of 4 *RPM.*, in the direction designated to give a forward thrust to the animal with reference to the platforms, speed of the perch was then gradually increased at 15 second intervals until the animal either fell off or jumped to one of the platforms If jumping occurred, the light was immediately turned off, the bird removed from the platform without difficulty[4] and placed in a dark compartment where it remained until the next trial. The period of two to five minutes' stay in the dark compartment followed every trial regardless of whether the bird had fallen or jumped. Records were kept of the perch speed at which each bird fell or jumped on each trial. The six animals used in this preliminary training procedure all developed a jumping habit in from 10 to 20 trials.

In establishing differential response to brightness or to pattern, *the above procedure was modified in two ways:, (a) stimulus cards* were fixed to the platforms, those on one platform constituting the positive or "rewarded" stimulus, those on the other serving as the negative or "punished" stimulus; (*b*) jumping to the negative platform resulted in the bird falling to the floor

---

[4] Since pigeons show little activity under dark conditions, turning off the light makes it possible to approach and remove the bird from the platform without producing incidental falls which would otherwise associate punishment with a correct response

RESULTS

Differential responses to black (positive) versus white (negative) stimulus cards were readily established in two normal pigeons. Ten trials were given daily and both animals made 10 errorless trials on the second day. However, two additional normal pigeons had not quite reached the criterion of 10 errorless trials at the end of 40 trials when training was discontinued.

Differential response to a black triangle (positive) versus a black circle (negative) was well established in a normal pigeon by the 40th trial. Likewise, a pigeon with a cerebellar lesion satisfied our criterion of differential response to a black square (positive) versus a black triangle (negative) by the 50th trial. Appropriate controls indicated that for each animal stimulus differences were effective in controlling the behavior.

DISCUSSION

In addition to its usefulness in problems of the discrimination type where incentive or motivational factors are traditionally treated as constants in the situation, our method should prove of use in the analysis of certain aspects of motivation. The fundamental task set for our animals permitted of alternative solutions in avoiding punishment (falling): *A*, improvement in ability to stay on the perch; *B*, improvement over chance in the selection of the correct platform on which to land. By reducing (but not to zero) the "availability" of Solution *A* by the preliminary training on the perch, all of our animals variously adopt Solution *B*. Under what conditions would such animals adopt Solution *A* rather than Solution *B*? By eliminating the preliminary training period, it would be possible to observe the stage in the development of the skill required to remain on the perch at which the animal will accept a problem for example in brightness or pattern discrimination as an alternative solution to a threatening situation. That more objective methods for the analysis of inter-level stresses are sadly needed by investigators of behavioral processes, is commonly recognized by experimentalists. The method outlined above constitutes, we believe, a promising approach to this problem.

SUMMARY

A jumping method is described for establishing differential responses to brightness and pattern in birds. In pigeons, differential response to

brightness could be established in as few as 20 trials and to pattern in approximately 50 trials. A use of the method in the study of motivation is suggested.

REFERENCES

1 GUNDLACH, R H The visual acuity of homing pigeons *J Comp Psychol.* 1933, **16**, 327-342.

1 HAMILTON, W F., & COLEMAN, T. B Trichromatic vision in the pigeon as illustrated by the spectral hue discrimination curve *J Comp Psychol*, 1933, **16**, 183-191

3 HAMILTON, W F & GOLDSTEIN, J L Visual acuity and accommodation in the pigeon *J. Comp. Psychol*, 1933, **15**, 193-197

4. KRECHEVSKY, I Brain mechanisms and brightness discrimination learning *J Comp Psychol*, 1936, **21**, 405-446

5 LASHLEY, K S. The mechanism of vision I A method for rapid analysis of pattern-vision in the rat *J Genet Psychol.*, 1930, **37**, 453-460

6 LAYMAN, J D. The avian visual system *Comp Psychol Monog*, 1936, **12**, (58), 1-36.

Otho S A Sprague Memorial Institute
and Division of Psychiatry
University of Chicago
Chicago, Illinois

Department of Psychology
Northwestern University
Evanston, Illinois

The Journal of Genetic Psychology, 1938, **52**, 233, 234

A NOTE ON THE HICCOUGH OF THE NEONATE*

Department of Psychology, The Ohio State University

ISABELLE F WAGNER

The hiccough, generally attributed to pressure of a full stomach upon the diaphragm, has frequently been observed in the infant shortly after birth However, no attempt has been made to proceed further and study the hiccough as an isolated behavior item. In the present study we have obtained pneumographic records of hiccough periods from 17 infants ranging in age from one to ten days

The hiccough period ranged in length from 35 seconds to 18 minutes 20 seconds, with a mean of 6 minutes 34 seconds. The total number of hiccoughs in a single period ranged from 11 to 337, with a mean of 96, while the mean interval between hiccoughs ranged from 2 6 seconds to 7.9 seconds, with a total mean of 4.5 seconds There was no apparent relationship between the total number of hiccoughs and the mean interval between hiccoughs in the same period

In general, the number of hiccoughs per minute decreased, with slight irregularities, throughout the period This was more obvious when the 17 records were combined. Since beyond the 12th minute only one record per minute was available in our data, the mean number of hiccoughs was obtained for only the first 11 minutes· 19, 17, 16, 14, 13, 12, 13, 10, 12, 14, 12 The number of cases was too small to obtain significant differences, but the data at least illustrated a dominant trend in the hiccough period.

There was no relationship between the age of the infant and the length of the hiccough period, the total number of hiccoughs in the period, or the mean interval between hiccoughs Nor were these last three factors related to the time at which the hiccough period occurred The infants were fed at 10, 1, and 4 during the day, and were usually available from half an hour after the time specified above to within half an hour before the next feeding. For the purposes of tabulation this two-hour interval was divided into eight

---

*Recommended for publication by F C Dockeray, and received in the Editorial Office on April 20, 1937

fifteen-minute periods numbered 1 to 8, and the data were ranked in terms of the period in which the hiccough *began* No parallel trends were revealed in the other three factors previously mentioned, but we found that most of the hiccoughs (14 cases) began during the first hour after feeding was completed, of these the major part (10 cases) began in the second half hour. One hiccough record was obtained in the fifth fifteen-minute period, one in the sixth, and one in the seventh. These findings support the popular belief that the hiccough nearly always occurs after feeding, but they indicate that the hiccough does not necessarily occur *immediately* after feeding

The accompanying illustration (Figure 1) gives a representative portion of a hiccough record This should be read from right to left The prolonged rise in the breathing curve at one point indicates a period of general bodily activity. The infant usually alternated, during a hiccough period, between conditions of general quiet and of vigorous activity, and occasionally cried, the eyes were generally wide open.

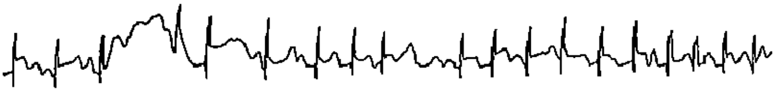


FIGURE 1
PNEUMOGRAPHIC RECORD OF HICCOUGH IN THE NEWBORN INFANT

The onset of the hiccough period was always sudden, but the termination of the period was less abrupt, the intervals between inspirations gradually lengthening, in an irregular fashion, until the hiccoughs had disappeared.

The hiccough varied in audibility from a barely audible inspiration to a loud, sharp sound. These loud sounds never accompanied all the hiccoughs in a given period, usually groups of them were interspersed with groups of relatively quiet inspirations.

A single isolated hiccough rarely occurred In records obtained from 200 infants, only five such instances were noted

Department of Psychology
The Ohio State University
Columbus, Ohio

The Journal of Genetic Psychology, 1938, 52, 235-239

THE SLEEPING POSTURE OF THE NEONATE*

Department of Psychology, The Ohio State University

ISABELLE F WAGNER

In a previously reported study dealing with the sleep of the newborn infant, supplementary records were kept of the posture changes of 95 of the infants used as subjects.

Since, in the previous study, conditions $A1_a$ and $A1_b$ [1] represented the stages of deepest sleep, the postures for these conditions are the only data considered here Conditions B (occasional stirs of body members) and C (infant generally active) by definition include fairly frequent movement, hence the variations in posture are likewise quite frequent.

The posture of the sleeping infant as determined from our data agrees only in part with the description given by Dennis (1) After carefully examining the literature, he concluded that the sleeping position of the newborn infant is as follows· mouth closed, legs flexed, fists closed; upper arms out straight from shoulder, forearms flexed at right angles so that they lie parallel to the head, fists often below chin as in embryonic life, however. Closure of eyes seems to be tacitly assumed in this description

From Table 1 we see that the mouth is closed most of the time, as Dennis found, but not always; in half as many cases the mouth may be slightly open, and in a few cases, at least half open In the $A1_a$ condition we find the mouth closed in 67 per cent of the cases; the mouth slightly open in 39 per cent, and the mouth open in 15 per cent. There is slightly greater possibility that the mouth will be closed in condition $A1_b$, for the per cents here are 79, 44, and 16 Frequently we find that the same infant may have the mouth closed in some $A1_a$ instances, for example, but slightly open in others

As for limb position, Dennis claims that the legs are flexed Table 1 shows us that while the legs most frequently are found

---

*Recommended for publication by F C Dockeray, and received in the Editorial Office on April 20, 1937

[1] $A1_a$. infant generally quiet, no eyelid or mouth movement, regular breathing. $A1_b$. same, except that breathing is irregular

TABLE 1

SLEEPING POSTURE

		Per cent of a given limb position			
		Both flexed	Both extended	One extended one flexed	No infants
Legs	A_{1_a}	68	45	51	82
	A_{1_b}	66	44	55	89
Arms	A_{1_a}	87	10	32	82
	A_{1_b}	84	15	37	99

Per cent of a given mouth position

	Mouth open	Mouth slightly open	Mouth closed	No. Infants
A_{1_a}	15	39	67	61
A_{1_b}	16	44	79	70

flexed (68% and 66% for A_{1_a} and A_{1_b} respectively), we also find one leg flexed and one extended in over 50 per cent of the cases (51% and 55%) and both legs extended nearly as often (45% and 44%). Here again, as with mouth position, the same infant may show a variety of specific postures in condition A_{1_a} and A_{1_b} in a given experimental period.

The upper arms, according to Dennis, go out straight from the shoulders, with forearms flexed at right angles so that they lie parallel to the head. In Table 1 we have called an arm "flexed" when the angle between the upper arm and forearm is 90° or less, and "extended" if the angle is greater than 90°; a similar distinction was made in the description of leg posture. The order of frequency of appearance of arm postures is the same as for leg posture. both flexed, one extended and one flexed, and both extended. However, the both-flexed posture predominates more in the arms than in the legs (87% and 84% in the arms), while there are smaller per cents of cases where both arms are extended as compared with both legs extended (10 and 15 for the arms).

The description, *"both arms flexed,"* is still not sufficiently specific It includes a variety of both symmetrical and asymmetrical positions of which the Dennis arm position is only one example. The asymmetrical positions are too varied to classify, but Part 2 of Table 2

TABLE 2
SLEEPING POSTURE
Per cent of a given arm position
(Part 1)

	One hand at head	Both hands at head	One hand on chest	Both hands on chest	No infants
A_{1_a}	22	4	12	2	82
A_{1_b}	24	3	20	2	99

(Part 2)

	(*a*)	(*b*)	(*c*)	(*d*)	No infants
A_{1_b}	45	18	6	10	82
A_{1_a}	43	20	6	17	99

indicates the most frequently found types of symmetrical arm position In the first, (*a*), upper arms are flexed at an angle of about 45°, with forearms parallel to the median axis of the body. In the second, (*b*), the forearms are rotated outward beyond a parallel position In the third, (*c*), we have the position described by Dennis In the fourth, (*d*), upper arms go out straight from the shoulders, but the forearms are flexed inward at an angle less than 90°. The order of frequency of appearance for both A_{1_a} and A_{1_b} is: (*a*), (*b*), (*d*), and (*c*). In other words, the arm posture described by Dennis is the exception rather than the rule, for it occurs in only 6 per cent (A_{1_a}) and 4 per cent (A_{1_b}) of the total number of cases for which posture was recorded

Dennis also stated that the fists often lie below the chin as in embryonic life To check this conclusion we classified the forearm positions thus: one hand at head (touching any portion of head or neck), both hands at head, one hand on chest (anywhere on torso below neck), and both hands on chest Since the hands rarely lay *at* the head or chest without actually touching it, no separate classifications were made for such positions In about one-fourth of the total number of cases one hand is touching the head (22% and 24%), somewhat less frequently one hand is touching the chest (12% and 20%), but in only a very small per cent of the cases are both hands touching the head (4 and 3) or the chest (2 and 2) Here again Dennis cited the exception rather than the rule.

Hand position is too variable to admit of convenient classification. Hands may be tightly closed, loosely closed with several fingers touching palm, loosely closed with only one or two fingers touching palm, almost closed, half open, open with fingers slightly flexed, or open with fingers spread They may be lying with palms up, on the side, occasionally with palms under, or halfway between two of the classifications The thumb may be tucked in under one or two fingers, or remain outside the flexed fingers, with or without the first thumb joint bent. Hand position is as frequently asymmetrical as symmetrical, for example, for one infant we record, "Right hand half open, on side, thumb out; left hand closed, on back, thumb in" However, cases of semi-closure or complete closure are more frequent than cases of hands wide open The latter cases generally are found immediately after a body jerk or sudden general stir, during which the fingers have been spread. Usually the fingers flex gradually thereafter to a position of at least semi-closure.

Head posture is not included in this discussion, for the newborn generally lies with the head on whichever side the experimenter places it. Only infrequently does the infant roll its head from one side to a median position or to the other side. Trunk position is likewise not discussed, for outside of a very few cases of a definite side position, the body lies in a dorsal position.

It may be objected that many of these posture variations are due to the constant stimulation of an infant. This can be only partially true, for some infants, however vigorous the body-jerk response in an A_{1_a} or A_{1_b} condition, relapse at once into the same posture as that preceding the response. Other posture changes are due to spontaneous body jerks not associated with an overt stimulus.

CONCLUSIONS

1. The sleeping newborn infant (conditions A_{1a} and A_{1b}) most frequently has legs and arms flexed and mouth closed.

2 There are far more variations in the sleeping posture of the newborn infant than have apparently been recognized heretofore. These variations occur not only from one infant to another but also in the same infant from one part of an experimental period to another

3 There are no consistent posture variations with age, at least within the first 10 days of an infant's life.

REFERENCE

1. DENNIS, W. A description and classification of the responses of the newborn infant *Psychol Bull*, 1934, **31**, 5-22

Department of Psychology
The Ohio State University
Columbus, Ohio

The Journal of Genetic Psychology, 1938, 52, 241-246.

A NOTE ON THE PERCEPTION OF LINEAR *GESTALTEN* IN THE RAT*

Swarthmore College[1]

I. KRECHEVSKY

The present note concerns itself with some data obtained from some preliminary experiments designed to investigate some of the neural implications of Koffka's "Trace-Hypothesis" (1) and the *findings of Kohler and Lauenstein relating to the effect of the* passage of time on the dynamic structure of such traces (Reported in Koffka, 1935).

In setting up the experimental situation for this problem, it is first necessary to demonstrate, for the rat (the subject chosen for the study since the brain-leison technique is to be used) certain fundamental relationships observed in the human subject in the field of visual perception.

The first such relationship to be determined is the effect of proximity of discontinuous stimulation (points) upon the organization of visual stimuli Proximity, as a principle of organization for visual stimuli, is too well-known to require elaborate discussion. A simple and common illustration will suffice. In the patterns in Figure 1, the dots form figures in which spontaneously the nearer ones unite to permit a perception of two horizontal linear distributions or groupings. Indeed, any other perception seems rather difficult to maintain. The common theory advanced for this phenomenon sup-

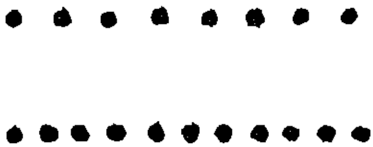


FIGURE 1

---

[1] The experimental work reported in this note was done in the Psychological Laboratories of the University of Chicago

*Received in the Editorial Office on May 14, 1937.

poses the "group formation as due to actual forces of attraction between the members of the group" (Koffka, 1935, p. 165). To determine, objectively, whether such a grouping does take place in the perceptual processes *of the rat* and whether the hypothesis suggested as an explanation for this grouping seems to be substantiated by the organism's behavior, was the specific problem of this first preliminary experiment.

PROCEDURE

Twenty-eight rats were trained to make a visual discrimination in the Lashley jumping-stand (2) Of these 28 animals only 16 finally mastered the discrimination-problem involved. (The animals were permitted a total of 250 trials on the problem before training was discontinued, and they were placed in the "unsuccessful" group) Of these 16 rats which met the criterion of learning (18 correct choices out of 20), half, or eight rats, were given the test which concerns our present problem specifically. This paper, then, will report on the data of these eight animals.

Specifically, the training procedure for these eight animals was as follows · (*a*) after a preliminary training period of seven days (during which the animals became accustomed to the requirements of the apparatus) the rats were started on the discrimination training between cards *"A"* and *"B,"* where card *"A"* was the positive stimulus (Figure 2). The animals were given 10 trials a day until the criterion mentioned above was reached. (*b*) On the very next day after the animal had met this criterion, he was given

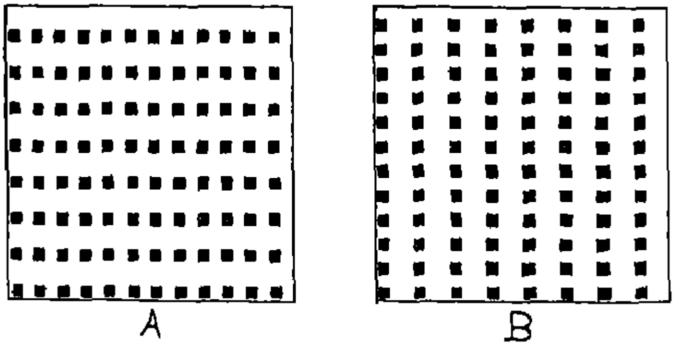


FIGURE 2

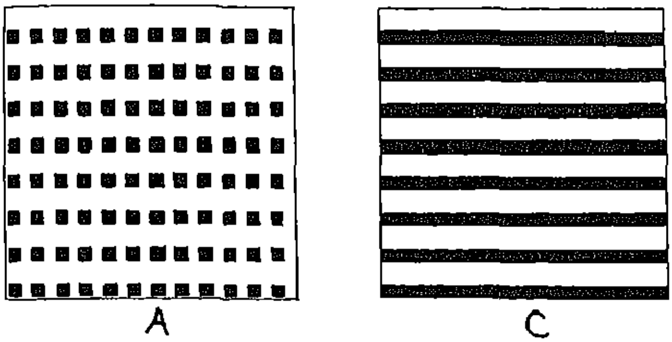


FIGURE 3

the test discrimination between cards *"A"* and *"C,"* (Figure 3). *In this test-situation a choice of either card was correct and was* rewarded equally. The test trials, also consisting of 10 per day, was continued for two days, or a total of 20 test-choices were given.

Data and Discussion

The arrangement of the stimuli on cards *"A"* and *"B"* is such, we believe, as to provide for an adequate test of our experimental question. Any black square on card *"A"* is nearer its neighbor to the right or to the left, than to its neighboring square above or below it. Just the reverse is true of the distribution of the squares on card *"B."*[2] In other words, if the factor of proximity operates to influence perception in the rat, the squares on card *"A"* should be grouped in a series of *horizontal* "groupings," and card *"B,"* in a series of *vertical* groupings. Regardless, however, of just what "grouping" the animals would finally achieve, the cards are so constructed, it seems to us, that unless some sort of grouping were made, *on the basis of the spatial distribution of the black squares,* no discrimination between the two cards would at all be possible. Both cards are exactly similar in all details except that one. Both have the same *area* of black against the same area of white, both have the same number of black squares, etc. Also, the possibility of the animals discriminating on the basis of extraneous cues was controlled by using different sets of cards every day.

---

[2] The dimensions for the stimuli-cards were as follows. For both cards the black squares were ¼ x ¼ inch. For Card *A* (Figures 2 and 3) the horizontal distance between the squares was ¼ inch, the vertical distance, ½ inch. For Card *B* (Figure 2) the reverse relationship held. The horizontal distance of separation was ½ inch, the vertical distance, ¼ inch.

Thus, it can be legitimately argued, we believe, that the very fact of learning in this situation is evidence for the hypothesis that the animals finally achieved a visual grouping of the stimuli distributed on the two cards *"A"* and *"B."* In Table 1 are presented the learning scores, in terms of errors made and trials required, for the eight animals used as subjects for this report. The problem appears to be a moderately difficult one requiring an average of 19 days, or 190 trials, for mastery. So much for our first point.

TABLE 1

ERROR AND TRIAL SCORES—ORIGINAL LEARNING

Rat No.	Total errors	Total trials
1	65	170
2	72	160
3	90	220
4	61	200
5	58	150
6	76	220
7	66	190
8	56	210
Total	544	1520
Average	68.0	190

The above data, however, do not prove that the grouping of cards *"A"* and *"B"* were such as would be required by the principle of proximity However, the data from the test situation seem to us to be most conclusive on this point. Observation of the pattern on card *"C"* (the card opposed to card *"A"* for the test trials) at once reveals that what we have here is a *completed,* or more perfect, form of the grouping which would result from the operation of the "forces of attraction" between the scattered squares on card *"A"* if the principle of proximity did apply. In other words, where we have the possibility of perceiving the squares on card *"A"* as consisting of a series of horizontal discontinuous groups, on card *"C"* we have an actual, and easily perceived, series of horizontal lines.

If, now, the animals, when presented with the choice between cards *"A"* and "B" should choose card *"A" in preference to "C,"* we can say nothing about the nature of the perceptual grouping which took place for the rat in the learning of the original problem, since these results might mean (*a*) that the animal actually did group the squares of *"A"* in horizontal "lines" but is still able to distinguish between *"A"* and *"C,"* or (*b*) that the animal did *not* group the squares into horizontal "lines" and therefore would naturally choose the card they had been trained to (card *"A"*) rather than a new card, *"C."* If, on the other hand, the animals showed *no preference for either card,* then we could say that the animals had grouped the squares on card *"A"* into horizontal groupings and were therefore reacting indiscriminately to either card *"A"* with its horizontal "lines" (phenomonologically) or to card *"C"* with *its* horizontal lines.

Finally, if the animals showed a *preference for card "C" as opposed to card "A,"* then we could say much more than that. We could, it seems to us, conclude that not only do the forces of attraction among the squares operate in accordance with the principle of proximity, but that these forces are of such strength and of such a nature as to make the organism prefer that stimulus-complex where the discontinuous members *in fact* coalesce as opposed to a stimulus-complex where the members are still, in some degree, discontinuous. All this may be conceived of, perhaps, as a specific instance of the operation of the well-known "Law of Pragnanz" in conjunction with the principle of proximity.

In Table 2 are presented the results of the *A-C* test trials. In

TABLE 2
TEST RECORDS

Rat No	Total "C" choices, 1st day	Total "C" choices	First choice
1	10	19	C
2	10	19	C
3	10	20	C
4	10	20	C
5	9	19	A
6	8	13	C
7	5	10	C
8	4	9	A
Total	66	129	6C;2A
Average	8 25	16 1	

the second column of the table are given the number of choices of the continuous lines (card "*C*") for the first day, and in the third column, the total number of card "*C*" choices for the two test days. From this table it is at once seen that despite the fact that for 190 trials (average number of trials, Table 1) the animals had jumped toward card "*A*," when presented with the choice between card "*A*" and a "new" card, "*C*," most of the animals showed a well-marked preference for card "*C*." As can be seen from the individual records, four of the eight animals chose card "*C*" 100 per cent of the time on the very first day of the test (rats Nos. 1, 2, 3, and 4, one animal made but one "*A*" reponse out of the ten trials (No 5); one animal, two "*A*" choices (No. 6) and the remaining two animals split their choices 50-50. Considering *their very first choice* in the test, column four of the table, much the same story is told. Six of the eight rats chose card "*C*" on their first test choice. No animal showed any appreciable preference for the card toward which they had been trained The total scores give the same results Of the 160 choices, 129 (or 80.6 per cent) were choices of card "*C*"

These data seem to us to indicate (*a*) that the rat tends to group discontinuous visual stimuli according to the principle of proximity very much after the fashion of the relationship observed in the perceptual behavior of the human being, and (*b*) that the "forces of attraction" between the members of such a visual group are of such a nature as to make the rat prefer the continuous *gestalt* over that of the discontinuous group *even though the original training be on the discontinuous stimulus-complex.*

REFERENCES

1 KOFFKA, K. Principles of gestalt psychology. New York: Harcourt, Brace, 1935

2. LASHLEY, K. S The mechanisms of vision *J Genet. Psychol*, 1930, **37**, 453-460

Swarthmore College
Swarthmore, Pennsylvania

APPARATUS

The Journal of Genetic Psychology, 1938, 52, 247-249

THE BLOCK PRINTING STICK AS A DEVICE FOR REGISTERING CHOICE*

Cambridge, Massachusetts

Marie L. H. Forbes

The drawing (Figure 1) shows a board with parallel slots (*c* and *c*′) through which tape passes from a spool with ratchet (*b*) to a printing cell (*d*), a board or tray (*a*) having geometrical holes in

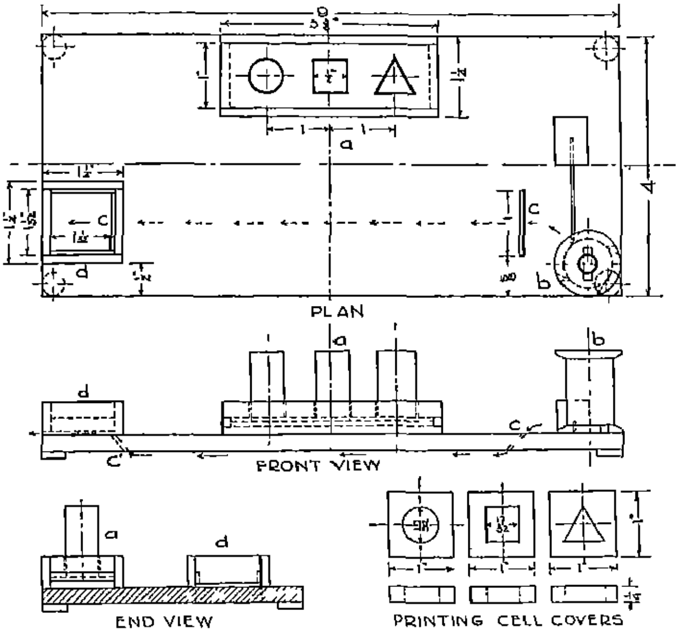


FIGURE 1

---

*Received in the Editorial Office on March 31, 1937.

which geometrical block printing sticks stand on an ink-pad, and a series of covers for the printing cell in each of which is cut a single geometrical hole.

The subject examines the cover of the printing cell which is already in position, selects the appropriate stick, and inserting it in the cover of the printing cell, brings it in contact with the tape. The tape may be pulled along at regular intervals of time without the cooperation of the subject.

A simpler printing board has two cells, each including a slot, the cell on the right containing a roll of tape which, unwinding counterclockwise, enters the slot without twisting.

The board with geometrical holes (*a*) somewhat resembles the three-hole form board of Gesell (3) used by the Kelloggs (4) and a tray with ink-pad (2) previously described.

As a variation in the device, the sticks may be alike in shape but unlike in color, or, as in the Dearborn color-form test (1), both color and form may be represented, printing cell covers being supplied to correspond. Perforated rubber, as indicated (Figure 2) in the diagram, may be mounted on the sticks for reproduction in black and white with a legend.

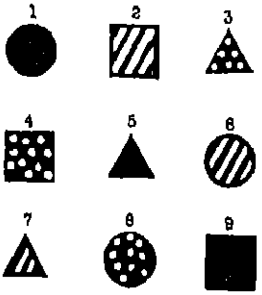


FIGURE 2

The writer gratefully acknowledges the examination of the materials by her former instructor, Dr. Walter F. Dearborn, Director of the Psycho-Educational Clinic, Harvard University.

REFERENCES

1. DEARBORN, W. F., ANDERSON, J E, & CHRISTIANSEN, A. O Form board and construction tests of mental ability. *J Educ. Psychol*, 1916, **7**, 445-447
2. FORBES, M. L. H. An arrangement of printing cubes with combined ink-pad and type-cell. *J. Genet. Psychol*, 1934, **44**, 437-439
3. GESELL, A The mental growth of the pre-school child New York. MacMillan, 1925
4. KELLOGG, W N, & KELLOGG, L A. The ape and the child. New York. McGraw-Hill, 1933

8½ Ash Street Place
Cambridge, Massachusetts

BOOKS

The Journal of Genetic Psychology, 1938, 52, 251-253.

LOIS BARCLAY MURPHY. *Social Behavior and Child Personality. An Exploratory Study of Some Roots of Sympathy.* New York, Morningside Heights: Columbia Univ Press, 1937. Pp. x+344

In this monograph, Dr. Murphy has given us a very detailed account of the reactions of nursery school children to manifestations of distress in others After a preliminary study during which the techniques of investigation were worked out, two groups of nursery school children totalling 39 subjects were observed for two-hour periods on two or three days of each week during an entire school year, making a total of 432 hours of observation, during which a *brief descriptive record was taken of all instances in which a definite* response of any kind to indications of distress by another child was observed Since the children in these schools were almost without exception the offspring of parents belonging to one or another of the learned professions and most of them were under four years of age, three other groups were also studied though less intensively. These comprised a pre-kindergarten group of four-year-old children, a group of day-nursery children from the lower socio-economic levels, and a group of hospital children (cripples, convalescents and children under observation) between the ages of three and seven years In addition to the observational data, teachers' ratings on a large number of personal-social characteristics were obtained, and 34 of the children were made the subjects of a controlled experiment in which concrete situations, stories and pictures designed to arouse sympathy were used as stimuli Information about the home and the family life was obtained by means of a standardized interview *with each parent and through day-by-day records kept by the parents* of six children.

The analysis of findings is far too elaborate and detailed to be adequately described in the space of a single review The usual statistical tables are supplemented by many diagrams, both theoretical and factual, and each point made is illustrated by a number of anecdotal citations of episodes taken from the formal records. Some readers may feel that the amount of space devoted to these illustrative anecdotes is excessive, since a rough calculation indicates that they make up at least a third of the entire volume and in many instances the points in question are of such a simple direct character that the citation of concrete examples seems like elaborating the obvious. On p. 117, for example, it is pointed out that children who are absorbed in play sometimes fail to notice the distress of others and so do not respond to it. This is understandable enough and hardly seems to require the 17 lines of rather trite anecdote that follow it. Here and elsewhere, the discussion would run much more smoothly if it were less frequently interrupted

Among the more significant findings may be mentioned the following· (*a*) Sympathetic responses tend to increase in frequency with advancing age throughout the nursery-school period and they undergo very distinct changes in pattern. Level of intelligence as measured by standard tests also seems positively related to sympathetic behavior though the correlation is not high. The social structure of the group with its elaborate interweaving of inter-child relationships is very important. The complicated nature of these relationships is nicely illustrated by a series of graphs in which the number of recorded contacts of each child with every other one is depicted by intersecting lines connecting their positions on the circumference of a circle. Some children show many social contacts, others few; some confine their social relations almost entirely to a small number of close friends, others extend them to include practically all the children in the group. Expressions of sympathy, therefore, are not wholly determined by the age and personality of the child who displays them or of the child who receives them but they vary with the immediate structure of the group situation and with the inter-personal relationships previously established. There is further indication that insofar as a "trait" in the sense of a relatively well-defined type of behavior that is characteristic of an individual may be said to exist, the evidence of the present study points to a more broadly displayed urge toward social contacts of all kinds rather than the more limited pattern defined as "sympathy" The children who most often showed sympathy at the distress of others were also found to display such aggressive and unsympathetic behavior as taking away another child's toys, pushing, pulling, or striking another child, or joining in the attack on one child by another, more frequently than did children in general. Tender and untender behavior seemed to be positively rather than negatively correlated, at least for the group in question. Nevertheless, the individual personality-pattern persisted in spite of these apparent contradictions for children who were in general aggressive tended to show their sympathy in a more forthright manner than those who were timid and shy. The latter were likely to weep with those who wept, or to offer verbal consolation while the more aggressive youngsters more often came to the defense of an injured child and used their fists against the aggressor.

The author wisely refrains from attempting to separate the closely interwoven strands of innate tendencies and environmental influences as determinants of sympathetic behavior. She points out, however, that each manifestation of sympathy is a response to an immediate situation as that situation appears to the child, and that the real or psychological stimulus is not external to the child but is a complex of all that the child is and all that he has experienced No experienced event is independent of the culture pattern in which the experiencing individual has been reared; no individual remains unchanged by the experiences he undergoes. Since this is so, variability rather than uniformity in behavior is to be expected. Yet this variability is not without order and consistency, for each experience carries meaning that shapes the personality as a whole and exerts a many-faceted effect upon behavior.

FLORENCE L. GOODENOUGH.

Institute of Child Welfare
University of Minnesota
Minneapolis, Minnesota

The Journal of Psychology

EDITED BY

Carl Murchison

The Journal of Psychology was founded at the request of many psychologists throughout the world for the purpose of establishing a medium in which publication shall be immediate and in which other publication experiments may be carried out. Beginning with January, 1936, the following methods and terms will be in force:

(1). The size of each volume will be 300-600 pages, and the subscription price will be $7 per volume, two volumes per year.

(2). The format of each volume will be such that each subscriber may take down the assembled volume into its component articles.

(3). Except through the fault of the author, the lag in publication should not be more than three or four weeks.

(4). The author will receive 200 reprints of his paper, and the subscriber will receive the accumulated publications bound into half volume units. The second half of each volume will contain the title page and table of contents for that volume. Each half volume unit will be mailed to the subscriber at its actual date.

(5). Each author will be charged the actual manufacturing costs of any zinc or half-tone cuts.

(6). The publication charges will be as follows:

4 pages or less	$12
8 " " "	16
12 " " "	24
16 " " "	32
20 " " "	40
24 " " "	48
28 " " "	56
32 " " "	64
36 " " "	72
40 " " "	80

These charges are tentative and will be adjusted as required by the subscriptions.

(7). The accumulated units will be mailed to subscribers quarterly: January, April, July, and October. The units may vary in size, to ensure immediate publication.

Beginning with January, 1936, eight hundred to twelve hundred pages (two volumes) annually. Per annum $14.00; per volume $7.00; per half volume unit $3.50. Complete sets at $7.00 per volume, plus transportation.

THE JOURNAL PRESS
2 Commercial Street
Provincetown, Massachusetts
U. S. A.

$7 00 per volume
Single numbers $4 00

QUARTERLY
Two volumes per year

June, 1938
Volume 52, Second Half

Founded by G. Stanley Hall in 1891

THE PEDAGOGICAL SEMINARY AND
JOURNAL OF GENETIC PSYCHOLOGY

Child Behavior, Animal Behavior, and Comparative Psychology

If this space should be unstamped, this is the regular library edition But if this space is stamped with a designating title, this is a special edition, sold under the restrictions of a bilateral contract, and may not be resold for a period of five years from date of publication

JUNE, 1938

(Manuscripts are printed in the order of final acceptance)

The function of punishment in learning . . . 257
By G C Drew

Personality characteristics of juvenile offenders in relation to degree of delinquency . 269
By Mervin A Durea

The constancy of mental test performance during the preschool period 285
By Marjorie Pyles Honzik

(OVER)

Copyright, 1938, by The Journal Press
Provincetown, Massachusetts
Entered as second-class matter August 3, 1897, at the post-office at Worcester, Mass, under the act of March 3, 1879
Reentered as second-class matter May 11, 1937, at the post-office at Provincetown, Mass, under the Act of March 3, 1879

Studies in mental development I. Performance on Gesell items at six months and its predictive value for performance on mental tests at two and three years 303
By Virginia Lafayette Nelson and T W Richards

Studies in mental development. II Analysis of abilities tested at the age of six months by the Gesell schedule . . 327
By T. W Richards and Virginia Lafayette Nelson

The use of pronouns by young children: A note on the development of self-awareness 333
By Florence L. Goodenough

The relation of mental stamina to parental protection . . 347
By John J B Morgan and Mary H Banker

The free word association of elementary school children II. Verbal responses 361
By William McGehee

Laterality in monkeys 375
By J. S Kounin

Performance of school children on the revised Stanford-Binet and the Kent E-G-Y test 395
By Arthur L Benton

The relation between the intelligence of mothers and of their children living in foster homes 401
By Donald Snygg

An investigation of reading retardation 407
By Arthur Lichtenstein

SHORT ARTICLES AND NOTES 425
Observations of dominance-subordination in cats (Charles Nelson Winslow) 425
The effect of early reaches on handedness in the rat: A preliminary study (Kenneth L Wentworth) . . 429
Achievement and intelligence examinations correlated with each other and with teacher's rankings (Harold A Kohn) . 433
A case of apparent imitation in a monkey (Josephine Ball) . . 439
The startle pattern in infants in response to non-auditory stimuli (William A Hunt, Frances M. Clarke, and Edna B Hunt) . 443

APPARATUS 447
A combined block printing cell and tape holder (Marie L H Forbes) 447

BOOKS 449

THESE JOURNALS WILL BUY BOOK REVIEWS . 449

BOOKS RECENTLY RECEIVED 453

The Journal of Genetic Psychology, 1938, 52, 257-267.

THE FUNCTION OF PUNISHMENT IN LEARNING*

Department of Physiological Psychology, Harvard University

G. C. Drew[1]

It is well known that learning is quicker when an animal is punished for the incorrect responses as well as rewarded for the correct, than when rewarded alone. This fact, described by the Law of Effect, and amenable to explanation in terms of the establishment of a positive conditioned reflex, accelerated by the contributory conditioned inhibition of "incorrect" activities, has been the basis of a body of theoretical discussion from which complicated conditioned reflex theories of learning have been deduced. Muenzinger (6-11) and his collaborators, however, have established beyond doubt that in accelerating learning, punishment for the correct response is as efficacious as punishment for the wrong. In the series of experiments cited, he has varied the position and the nature of the "punishment," and has found that electric shock for the correct response, electric shock after the point of choice (i.e., for both right and wrong responses), and jumping a gap in the floor after the point of choice all produce an acceleration in learning which is insignificantly different from that produced by electric shock for the wrong response. Placed before the point of choice, however, both electric shock and a gap in the floor slightly retard learning.

The present writer has obtained somewhat similar results under slightly different conditions. These experiments were performed, in the one instance, on a short wide runway (1), and in the other, on a parallel alley maze (2), and were designed to test the effect of the introduction of punishment, coincident with the reward, on the performance of an established habit. In the runway experiment a bell was used as a "punishment," being sounded during the time the rat was eating in the food-box. The value of the bell as a punishment was tested on a separate group by removing the food after learning was completed, and sounding the bell during the time the

---

*Accepted for publication by K. S. Lashley of the Editorial Board, and received in the Editorial Office on July 28, 1937.

[1] The writer wishes to acknowledge his gratitude to Prof. K. S. Lashley, whose advice and criticism were most valuable.

animals should have been eating. Under these conditions the bell disrupted the habit in three or four trials, which was far quicker than the disruption following removal of the food alone. Electric shock was substituted for the bell in the maze experiment, and was administered directly through the food. The experiments differed from Muenzinger's, therefore, in that punishment was not present from the beginning of learning but was introduced only after the habit was fairly well established, and that it was much more closely associated with the reward than in his experiments. In both problems the efficiency of the performance was increased following the introduction of the punishment.

It was suggested as a result of these experiments, that the increase in speed of performance and in the stability of the habit might be due to some factor of emphasis, such as that suggested by Tolman in his Law of Emphasis (12). Muenzinger, on the other hand, is led from his results to postulate "an increase of sensitivity to cues to be discriminated, . . due to its (the shock's) startling or shaking-up effect." Later, his failure to get acceleration with a buzzer, which did startle the animals, compared with the acceleration produced when they were required to jump a gap in the floor, a reaction which he reports as having no emotional significance, led him to explain the acceleration as being probably due to an increase in sensitivity to cues to be discriminated, but omitting any suggestions as to the nature of the process resulting in such an increase in sensitivity. In this form, the "increased sensitivity to cues" does not exclude an explanation such as that suggested by Tolman. It is, in fact, capable of being interpreted as a restatement of the Law of Emphasis.

The conception of "emphasis" as the determining factor is interpretable in two ways It can mean either emphasis of some specific parts of the situation, as for example, by emphasizing the light as against the dark door in a discrimination problem, making it easier for the animal to "see" the correct "hypothesis," or a general heightening of the animal's sensitivity by increasing his alertness or cautiousness or some such general factor. The evidence bearing directly on these problems does not give any indication as to which of these two possible interpretations is more valuable as an explanation of the accelerating effect of shock.

The relevant evidence from other work on learning problems is conflicting Lashley, in a personal communication, has stated his impression, gathered from a long acquaintance with puzzle boxes, that learning is quicker in such a situation if the door makes a slight noise on opening when the animal manipulates the correct objects Such a view favors the specific nature of the emphasis Similarly, the conditioned reflex theories explaining quicker learning with punishment for wrong responses in terms of a quicker inhibition of those wrong responses can be interpreted as supporting this view On the other hand, experiments such as that of Morey (5) in which the presence of a loud sound, whose source is at some distance from the apparatus, results in a quicker performance of the habit, suggest a general factor.

The present experiment was an attempt to obtain evidence which would throw some light on the process responsible for the acceleration in learning, by subjecting the concept of emphasis to experimental analysis. Shock was directly associated with those features of the situation which presumably are most important in the learning of a discrimination habit, i e, the right and wrong stimuli and the food. It seems logical to suppose that if emphasis of a specific nature is responsible there will be some differentiation between groups receiving shock in these different places. This should show up particularly in the differentiation between the shock-in-the-food group and the groups shocked on either of the doors. If, on the other hand, no significant difference is found, any explanation in terms of specific emphasis must be open to serious doubt, since it seems unlikely that emphasis on any one of these factors will have exactly equivalent functional significance with emphasis on any other one In the event of there being no differentiation between these groups it seems likely that some general behavioral factor, such as increased excitability or increased cautiousness must be held responsible.

Apparatus and Procedure

Forty-three rats were trained, light positive, on a brightness discrimination problem. The ground plan of the apparatus used (Figure 1) resembled Lashley's modification of the Yerkes' discrimination box (4). It consisted of a wide choice compartment from which two parallel alleys led to the stimulus doors, which the animals were forced to open to reach the food compartments The doors were diffusing glass screens, illuminated from behind, and, as in

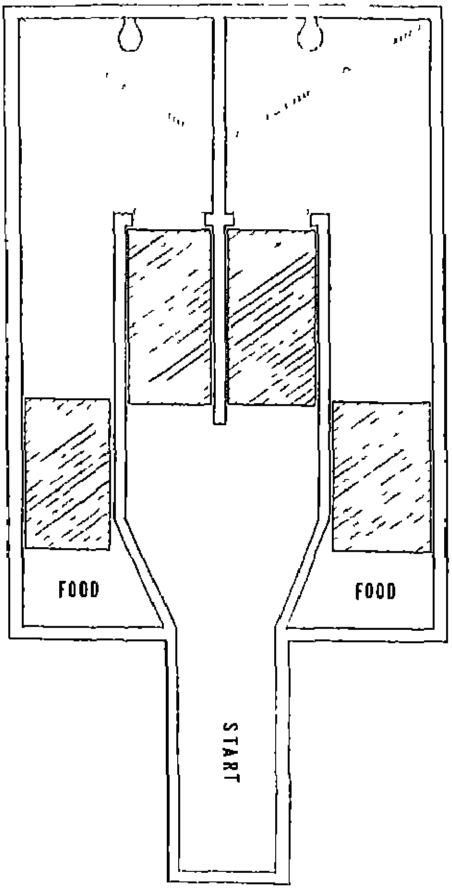


FIGURE 1

Lashley's apparatus, were the actual stimulus objects to be discriminated The conventional method of giving electric shock in this type of apparatus is to place grids in the approaches to the doors. For the success of this experiment it was necessary that the shock should have as direct an association with the stimulus objects as possible. As Köhler (3) has pointed out there is "only a very loose connection in space and time" between the shock and the negative stimulus in the conventional use of the grid. For a study designed to show *how* shock exercises its accelerating effect, the possibility that the animal must first learn the connection between the pain in its legs and the stimuli to be discriminated, is an undesirable complication which might produce less clear cut results, rendering their interpretation more difficult In this experiment, therefore, the apparatus was wired so that the animals could be shocked actually by the doors and food, receiving the shock not on their feet only, but primarily on their noses, as they pushed against the doors, or dipped into the dish for food

The principle adopted both for the doors and the food was to force the rats to complete a circuit between these objects and the floor of the apparatus To electrify the doors, a fine bronze mesh was fixed over the glass screens and was wired to one pole of the circuit Brass plates were then screwed in the parallel alleys leading to the doors, and were in turn wired to the other pole The rat was forced to stand on the brass plates in order to open the doors, and, in pushing against them, completed the circuit. Similarly, the food, moist bread and milk, was placed in a non-conducting dish and was wired, through a small platinum wire in the bottom of the dish, to the same pole as the doors The dish was placed at the end of the narrow food compartment, and the approach to it covered with a brass plate, wired as for the doors To eat, the rat was forced to stand on the plate and dip his head in the bread and milk, again completing the circuit. The moistness of the food, and its amount relative to the height of the dish were kept approximately constant, in an attempt to control the current density. Owing to current density, then, the rat should feel the shock on the part with which he touched the doors and food. Observation of the animal's behavior indicated that this was indeed the case Each door and the food was wired separately and was controlled by an independent switch in front of the experimenter The intensity of the shock given was approximately 15 milliamperes, alternating current.

To prevent the animals learning to push against the doors with their shoulders or some other part well covered with hair, and thus greatly diminish the intensity of the shock, the doors were hinged at the bottom so that with the pressure of the rat's body against them they swung downwards, forcing him to walk across them to reach the food. The doors were carefully counter-balanced so that they moved to a very slight touch, and would swing gently back into place as soon as the rats had crossed

The animals were divided into four groups, the control having 10 animals and the other three 11 each. Of the three experimental groups, one received shock on the incorrect door, one on the correct, and one in the food. After the preliminary training, in which they were taught to open the doors and to eat in the food compartments, the rats were given 1 trial on the first day and 2 on the second, without shock, then 2 trials on the third day with shock, and 5 trials a day thereafter until the eleventh day, when the number of trials daily was increased to 10. The criterion of error used was actually touching the incorrect door.

RESULTS

All the animals were given 80 trials. The results are shown graphically in Figure 2, which gives the mean error curves for the

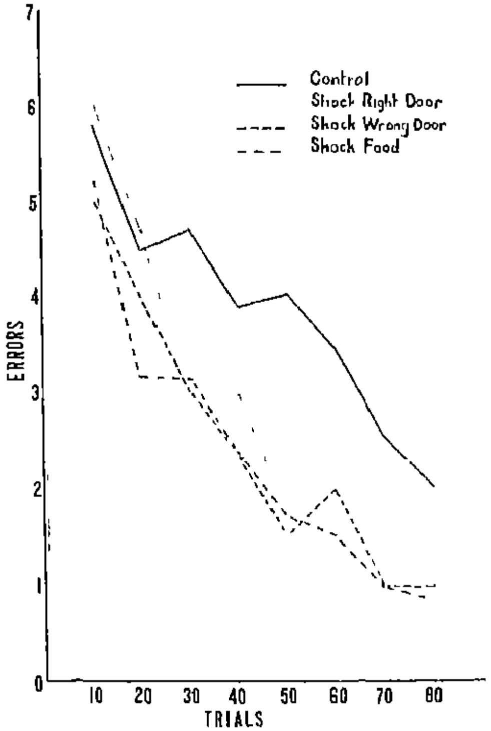


FIGURE 2

four groups. It will be seen that in all three shock groups there is a well marked acceleration in learning, as compared with the rate of learning of the no-shock group. The differences between the shock groups and the control are statistically significant, the critical ratios of the differences ranging from 3.54 to 5.50. These critical ratios, with the insignificant critical ratios of the differences between the various shock groups, are given in Table 1.

TABLE 1

THE FIRST NAMED OF EACH PAIR OF GROUPS MADE FEWER ERRORS

Groups	Critical ratios
Shock Food and Control	5 50
Shock Right Door and Control	3 70
Shock Wrong Door and Control	3 54
Shock Food and Shock Right Door	1 10
Shock Food and Shock Wrong Door	0.32
Shock Wrong Door and Shock Right Door	*0 52*

The curves for the three shock groups are almost identical after the first 20 trials The curve for the shock-in-the-food group is somewhat smoother than those for the other shock groups, but the difference is not sufficiently great to merit any theoretical consideration During the first 20 trials differences may be found between these three groups, but they are not large enough to be statistically significant. They are probably due to certain qualitative differences in the reactions of the rats to the various positions of the shock. Up to the twentieth trial, the group receiving shock on the correct door is slightly worse than the control group This is due to the obstruction-like action of the shock on the door. Characteristically, the animals in this group would go to the correct door as frequently as those in the other shock groups, but on receiving the shock would retrace to the incorrect door, and would spend some time trying to open it before returning to the correct one. The choice was counted an error if the animal pushed against the incorrect door, irrespective of which he tried first This criterion of error has made the error scores for the rats in this group abnormally high during this period when they were learning to overcome their reluctance to open the door in spite of the shock This obstructive effect resulted in slower and more cautious reactions than those of rats in the other groups. During these early trials "vicarious trial and error," as Muenzinger has named the phenomenon of hesitation at the point of choice, with head movements in the direction of the two stimuli, appeared much more frequently with this group than with the others.

The behavior of the animals shocked in the food did not show the obstructive effect of the shock, with the exception of some hesitation about eating on the early trials In the apparatus these animals showed marked signs of excitement, rushing to either stimulus door without appearing to make any choice. The difference between this group and the control group, apparent even on these early trials, and the rapidity with which they obtained better than chance scores indicate that they were in fact making choices, in spite of their apparent random reactions. Unlike the group shocked on the correct door, if they went first to the incorrect door, they retraced hurriedly and opened the correct one without any hesitation, whereas the shock-on-the-correct-door group tended to struggle with the incorrect door for some time before retracing to the correct. With the shock-wrong group, on the other hand, there was no characteristic behavior apparent, the group behaving in a manner similar to that of the control group.

It must be stressed that these qualitative differences were not subjected to systematic observation and that they represent merely the experimenter's impressions of the various groups.

Discussion

It might have been predicted from Muenzinger's results that shock on the wrong and on the correct doors would produce a fairly uniform acceleration in learning compared with the learning of a group with no shock. Apart from the writer's previous experiments, which from the smallness of the groups and the difference in technique have only a suggestive value here, there is no evidence on which one could predict the effect on learning of the shock in the food The fact that the acceleration produced by such a shock is practically indistinguishable from that produced by shock for the correct and incorrect responses respectively, affords striking confirmation of Muenzinger's contention that shock anywhere after the point of choice will produce an acceleration in learning.

It was suggested at the beginning of this paper that the conception of emphasis as the factor producing the acceleration in learning might be interpreted in either a specific or a general sense, and it was assumed that if the emphasis is of a specific nature there would be some differentiation between the effects of shock in different parts of the stimulus situation. This assumption was made since it is conceivable that, in a light-dark discrimination, emphasis placed on either of the two stimuli would be equally efficacious in accelerating learning, but it is scarcely reasonable to suppose that emphasis on the reward would have an exactly comparable effect It has been shown, however, that within the limits of this experiment, the locus of the shock is unimportant, the effect of the shock being equivalent for all three groups This makes it seem unlikely that emphasis of a specific nature is responsible for the acceleration produced

Since certain qualitative differences were observed in the behavior of the three shock groups, it is possible that under other conditions, involving a longer learning process and a greater variety of reactions than are possible with this apparatus, differences would be found which would be amenable to quantitative analysis. Such qualitative differences as were obtained can be used only to support quantitative differences or indicate new lines of attack They cannot, justifiably, be used to indicate differences where no quantitative differences can be found As careful analysis of the records failed to reveal any supporting quantitative differences in this experiment, it would seem best to conclude at present either that they were not true differences, or that they were irrelevant to the learning process proper.

It may be assumed, then, that some general factor is responsible for the acceleration. It was pointed out earlier in this paper that the term "emphasis" is capable of carrying a general meaning, as well as its more precise specific implications However, the term is most frequently used to imply emphasis of a specific nature, and the possibility of two distinct interpretations may lead to confusion Hence it is suggested that the term *"emphasis"* be restricted to its more usual specific meaning and that the factor operative in such a situation as that described in this paper be referred to merely as a "general factor" until further knowledge suggests a more adequate terminology

What this general factor may be is, at the present, merely a matter of speculation. From the nature of the stimuli used—electric shocks, buzzers, jumping gaps, it is plausible to assume that increased excitability, or increased excitation is the factor responsible. The results of Muenzinger's shock-before-choice experiments, however, make this theory unlikely. Other possible factors, such as increased cautiousness, increased attention, or increased alertness likewise have no experimental evidence to prove or disprove them. At present, we are justified in saying only that the factor responsible for the acceleration is probably a general rather than a specific one.[2]

SUMMARY

Forty-three rats were trained, light positive, on a light-dark discrimination problem, in an attempt to analyze the concept of emphasis as the factor responsible for the acceleration in learning produced by the addition of punishment to the hunger-food motivation.

The animals were divided into four groups, a control no-shock group and groups shocked on the incorrect door, the correct door, and the food respectively. In all cases shock was administered on the doors and food themselves, and not by grids.

The results show a significant acceleration in learning for all three shock groups compared with the control group Moreover, the acceleration is uniform for these shock groups.

These results confirm Muenzinger's contention that shock anywhere after the point of choice is equally efficacious in accelerating learning. From this it is concluded that emphasis, interpreted in a specific sense, cannot be the factor responsible. It is suggested, rather, that a general factor seems to be operative The nature of this general factor is at present a matter of speculation

REFERENCES

1 DREW, G C. The effects of a mixed incentive on the behavior of rats *Brit J Psychol*, 1935, **26**, 120-134.

2 ——— The growth of a conflict between an established habit and an antagonistic activity (In press)

3. KOHLER, W Intelligence in Apes *Psychologies of 1925* (C. Murchison, Ed) Worcester, Mass.: Clark Univ. Press, 1926 Pp 412.

---

[2] Since this paper was written Muenzinger has published evidence which leads him to suppose that shock operates by enforcing a delay at the point of choice This supports the contention of the present paper that shock has a general rather than a specific effect.

4. Lashley, K S. The mechanism of vision XII. Nervous structure concerned in the acquisition and retention of habits based on reactions to light *Comp Psychol. Monog*, 1935, **11**, 43-79

5 Morey, R H. The swimming speed of rats as a function of the presence or absence of sound *J Comp. Psychol*, 1934, **17**, 329-354

6 Muenzinger, K F Motivation in learning I Electric shock for correct response in the visual discrimination habit *J Comp Psychol.*, 1934, **17**, 267-277

7 Muenzinger, K. F, & Newcomb, H Motivation in learning III. A bell signal compared with electric shock for right and wrong responses in the visual discrimination habit *J. Comp Psychol*, 1935, **20**, 85-93.

8. ——— Motivation in learning V The relative effectiveness of jumping a gap and crossing an electric grid in a visual discrimination habit *J. Comp Psychol*, 1936, **21**, 95-104.

9 Muenzinger, K F, & Fletcher, F M Motivation in learning VI Escape from electric shock compared with hunger-food tension in the visual discrimination habit. *J Comp Psychol*, 1936, **22**, 79-91

10 ——— Motivation in learning VII. The effect of an enforced delay at the point of choice in the visual discrimination habit. *J. Comp Psychol.*, 1937, **23**, 383-392

11 Muenzinger, K F, & Wood, A.. Motivation in learning IV The function of punishment as determined by its temporal relation to the act of choice in the visual discrimination habit. *J Comp Psychol.*, 1933, **20**, 95-106

12 Tolman, E C Purposive Behavior in Animals and Men New York Century, 1932.

Department of Physiological Psychology
Harvard University
Cambridge, Massachusetts

The Journal of Genetic Psychology, 1938, 52, 269-283

PERSONALITY CHARACTERISTICS OF JUVENILE OFFENDERS IN RELATION TO DEGREE OF DELINQUENCY*

Department of Psychology, Ohio State University

MERVIN A DUREA

INTRODUCTION

Knowledge of factors constituting the personality of juvenile delinquents is of utmost importance for several reasons Differential diagnosis should be based upon demonstrated facts of personality rather than inferences An accurate picture of the delinquent's personality is necessary before a reliable prognosis of subsequent behavior and adjustment can be made How the individual delinquent will conduct himself in the future depends on the nature of present tendencies The institution of proper therapeutic procedures is governed in large part by an understanding of personality characteristics of the juvenile offender Whether therapeusis involves an alteration in the social situation or necessitates some form of personal re-education, such cues as are derived from an analysis of the individual's personality indicate the ultimate course to be followed.

PROBLEM

Discovery of traits of personality which effectively differentiate between subjects with the lowest and highest degree of delinquency is the fundamental objective of the present study Factors of personality under consideration include circumstances judged as wrong; fears, anxieties, and worries; objects of interest, and qualities admired in others Contrasted groups have been employed as a basis for the analysis.

PROCEDURE

Results obtained from the *Interest-Attitude Tests,* a technique devised by Pressey (6), were used in this investigation This instrument consists of four tests, each containing 90 items, including

---

*Accepted for publication by Buford Johnson of the Editorial Board, and received in the Editorial Office on August 9, 1937

things considered wrong (Test I); anxieties, fears, worries (Test II); likes and interests (Test III); and kinds of people liked or admired (Test IV). The subject is instructed to respond discriminatingly to the 90 items embodied in each test Thus, from Test I the first five items are: *accidents, fighting, ignorance, talking back,* and *crying.* In relation to these items and others similar the subject indicates by *one* cross (X) everything considered *wrong* and by *two* crosses (XX) items suggesting circumstances regarded as *very wrong.* If a given item is single-crossed by 42 and double-crossed by 25 subjects from a group of 80, the total number of responses would be 92, and the number of responses per 100 cases 115. Norms are available for the separate tests by sex and grade (sixth grade to fourth year college) in terms of the number of responses to each item per 100 cases.[1] Hereinafter, therefore, the designation *times-in-100* has been adopted as the differential unit for comparative purposes.

Contrasting groups, representing the least and most serious aspects of juvenile delinquency were obtained by a method reported by the writer (4). Essentially it consists of the utilization of three operations the numerical values of which, equally weighted and compounded into a total, quantitatively express varying degrees of delinquency. The three operations are: length of time a subject has been an offender (duration); total number of times a subject has appeared in Juvenile Court (frequency), and the total of weighted values assigned to the different forms of delinquent behavior committed by the subject (scale values).[2] Operational values were reduced to standard scores, these were totalled, each value receiving equal weight, the sum constituting a Delinquency Index (*DI*).

From a frequency distribution of 316 *DI*s computed on boys resident in an institution for juvenile delinquents, 64 subjects having the lowest *DI* and the same number having the highest *DI* were selected. The basic sample and selected contrasted groups were of the white race. Life ages ranged from 14 years, 0 months to 17 years, 11 months. The sample appeared to be fairly representative of delinquent boys in general Life age was ignored in the selection of contrasting groups.

---

[1] Published by The Psychological Corporation, New York, N Y

[2] Details for classifying different forms of delinquent conduct and the method for deriving weights for each type of offense have been reported elsewhere (3).

Delinquency Indexes of the 316 juvenile offenders ranged approximately from 105 to 250, the median being 148, the mean 150. Lowest 64 *DI*s ranged from about 105 to 130, the median being 122; highest from about 170 to 250, the median being 181 The groups contrasted as to Delinquency Index included about 40 per cent of the total group of delinquents.

The first step in the analysis was to ascertain the frequency with which subjects having lowest *DI*s and those with highest *DI*s responded to each item Resulting frequencies were reduced to the differential unit, times-in-100 Item-by-item differences were computed between the contrasted groups based on the proportionate times-in-100 responses were made by each group If subjects in the group of high *DI*s responded more times-in-100 than subjects with low *DI*s the difference value was assigned a plus sign If the converse were true the difference value was denoted with a minus sign Thus, Item 1 (accidents) in Test I had a frequency of 44 or 69 times-in-100 for high *DI*s and 38 or 59 times-in-100 for the low *DI*s, the difference value being +10. Item 10 (atheist) of the same test had a frequency of 11 or 17 times-in-100 for high *DI*s and for low *DI*s a frequency of 16 or 25 times-in-100, the difference value being —8. Computations similar to those described were performed for each of the 90 items constituting each test.

To select items which were of unquestionable significance in distinguishing between least and most serious delinquent subjects the following procedure was adopted. difference values for each series of 90 items of the four *Interest-Attitude Tests* were cast into frequency distributions Plus and minus designations were ignored since the primary purpose of signs was to indicate *direction* of differences. For each distribution of difference values Q_3 or the 75th percentile was calculated The value of Q_3 for the difference values of Test I was 16; for Test II, 14; for Test III, 14, and for Test IV, 14 Table 1 makes clear the foregoing procedure

Items the difference values of which equalled or exceeded the 75-percentile of the series of differences in question were regarded as *basically significant* in differentiating between the groups of high

TABLE 1
FREQUENCY DISTRIBUTION OF DIFFERENCE VALUES FROM EACH TEST OF THE INTEREST-ATTITUDE TESTS

Class Interval	Test I F	Test II F	Test III F	Test IV F
0-4	21	28	31	26
5-9	29	27	26	25
10-14	15	17	14	22
15-19	13	11	6	13
20-24	3	4	5	4
25-29	4	1	5	0
30-34	0	2	2	0
35-39	4	0	0	0
40-44	0	0	1	0
45-59	1	0	0	0
Total	90	90	90	90
Md.	9.1	8.1	7.7	8.8
Q_3	16.0	13.7	13.8	13.8

TABLE 2
TECHNIQUE FOR SELECTING ITEMS DIFFERENTIATING BETWEEN LEAST AND MOST SERIOUS JUVENILE DELINQUENTS

Item No.	Item	Times-in-100 (High *DI*)	Times-in-100 (Low *DI*)	Difference Value
1	accidents	69	59	+10
4	talking back	88	123	—35
10	atheist	17	25	— 8
20	bribery	81	73	+ 8
80	lawlessness	72	100	—28

and low *DI*s. Table 2 has been arranged to illustrate the operation of this principle.[8]

Table 2 shows five items from Test I. Applying the criterion just expressed, it is seen that Item 4 (talking back) and Item 80 (lawlessness) qualify as basically significant items, their difference values being well in excess of a Q_3 of 16.

Evidently in the light of procedures which have been adopted, four series of significant items from the *Interest-Attitude Tests* will be the outgrowth. Since Q_3 of each series of difference values is the value-point which the difference value of the single item must equal or exceed in order to be regarded as basically significant, such lists of differentiating items will contain in the neighborhood of 20 items each, all Q_3's being computed on the basis of 90 difference values As a matter of fact basic lists ranged from 19 to 26 items.

---

[8] Selection of Q_3 as the value which difference values of items must equal or exceed in order to be considered as basically significant was entirely empirical. Each Q_3, however, was an expression of the quantitative tendencies of the data. It was assumed that the sensitivity of items in differentiating least from most serious delinquents would become greater and greater in proportion as difference values were in excess of the mode.

Now presented is the important question of settling on a principle by means of which *maximally effective* items are to be selected from the more comprehensive series of significant items As a practical solution of the problem, maximally effective items are defined as those which exceed Q_3 by 10 or more, this assumption being applied in each of the four series of significant items

STATISTICAL ANALYSIS

From comprehensive tabulations of responses made by the groups contrasted, two facts issue that are of fundamental interest· first, items which from each of the four *Interest-Attitude Tests* are broadly significant in differentiating between low-high *DI*s, second, items which from each significant series are maximally effective as differentiae To clarify these two aspects of analysis is the purpose of Table 3.

Table 3 shows basically significant items from each test Maximally effective items are italicized Immediately after each item is its difference value together with directional plus or minus sign.

From the standpoint of total number of significant items shown in Table 3 several interesting facts are revealed. Of 360 items embodied in four tests of the *Interest-Attitude Tests* 92, or 26 per cent, are basically significant Of the 92 items 50, or 54 per cent, are plus items and 42, or 46 per cent, are minus items. Of the 42 minus items 37 appear in Tests I and IV. Of the 50 items denoted by a plus 40 appear in Tests II and III. This means, on the one hand, that with reference to circumstances considered to be wrong and qualities admired in other people, there is a definite weighting in the direction of least serious juvenile offenders. On the other hand, anxiety tendencies and interests are weighted toward the most serious delinquent group.

Scrutiny of significant items of the separate tests further clarifies the manner in which least-most serious delinquents are differentiated. In Test I there are 23 significant items, only one of which, *yellow-*

TABLE 3

BASICALLY SIGNIFICANT AND MAXIMALLY EFFECTIVE ITEMS FROM INTEREST-ATTITUDE TESTS

		TEST I			
fighting	—36	petting	—17	gang	—17
talking back	—35	arguing	—17	*playing cards*	—28
crying	—20	poker	—20	playing hookey	—16
pawning jewelry	—22	*tobacco*	—38	being queer	—17
carrying a revolver	—18	bashfulness	—16	*quarreling*	—36
teasing someone	—19	*pool rooms*	—27	shouting	—16
smoking	—45	betting	—17	*lawlessness*	—28
shooting craps	—26			yellowness	+16
		TEST II			
detective	+19	coffin	+20	feebleness	+14
murder	+22	being unlucky	—14	thieves	+15
poison	+27	craziness	+14	*cyclones*	+31
choking	+14	headache	+14	funeral	+19
skeleton	+14	*jail*	+30	robber	+17
grave	+14	dreams	+15	sickness	—16
suffocating	+18	sins	+19	*operation*	+24
lightning	+14	disease	+19	wrecks	+16
crimes	+21			self-consciousness	+19
		TEST III			
cartoonist	+21	*baseball*	+30	*roller skating*	+28
movie star	+28	animal trainer	+19	children	—20
comedies	+28	*card parties*	+28	coffee	+17
soldiers	+26	poker	+15	Sunday School	—15
fiction	+21	sailors	+20	*shooting*	+42
gym work	+15	dominoes	—17	wireless operator	+20
		baseball player	+32		
		TEST IV			
courageous	—17	competent	—17	sincere	+14
progressive	—19	having initiative	—20	innocent	—16
discreet	—19	dependable	+19	*rich*	+24
efficient	—14	expert	+17	up-to-date	+14
generous	—21	able	—16	humorous	—14
fashionable	—14	inventive	+19	witty	+14
patient	—17	wealthy	+16	busy	—21
lively	+15	well-informed	—14	accomodating	—15

N.B. To facilitate interpretation of items in Table 3 it should be recalled that Q_3's for the four series of difference values resulting from Tests I to IV are Test I = 16; Test II = 14; Test III = 14; Test IV = 14.

ness, bears a plus notation. Twenty-six items in Test II are significant, two of which, *being unlucky* and *sickness,* have minus designations. Of 19 significant items in Test III, three have directional signs of minus The three minus item are *dominoes, children,* and *Sunday School* Significant items in Test IV total 24, of which 9 are plus and 15 are minus Thus far it appears reasonable to conclude that from one angle tendencies to fear and anxiety and, from another, interests, best differentiate subjects with a high degree of delinquency Further, blame-worthy circumstances and characteristics admired in other people are most differential for least serious delinquents.

From items which are broad in their significance attention now narrows to special items which are maximally effective in differentiating between groups contrasted in terms of degree of delinquency, namely, low-high *DI*s Items have been categorized as maximally effective if their difference values exceed the given 75-percentile by 10 or more points As stated, maximally effective items in Table 3 are shown in italics. A fact which will be noticed at once with reference to maximally effective items is that considerable variation exists in the extent of their differences from respective Q_3's This has suggested the possibility of arranging each array of maximally effective items in the form of a hierarchy For example, Item 15 (smoking) of Test I has a difference value of —45, and the difference value of Item 19 (shooting craps) of the same test is —26, the highest and lowest difference values of the nine maximally effective items from Test I. Therefore, in the hierarchy *smoking* will rank first and *shooting craps* ninth Pursuant to the principle described, maximally effective items from Tests I, II and III are shown hierarchically arranged Only a single item from Test IV fulfilled requirements for maximal effectiveness

Nine maximally effective items from Test I follow, each denoted by its proper directional sign and difference value

> *Smoking* —45, *tobacco* —38, *quarreling* —36, *fighting* —36, *talking back* —35, *playing cards* —28, *lawlessness* —28, *pool rooms* —27, *shooting craps* —26

From Test II four items meet the standard for maximal effectiveness in differentiating least-most serious delinquents

> *Cyclones* +31, *jail* +30, *poison* +27, *operation* +24

Maximally effective items from Test III are eight in number

> *Shooting* +42, *baseball player* +32, *baseball* +30, *movie star* +28, *comedies* +28, *card parties* + 28, *roller skating* +28, *soldiers* +26

Although Test IV as seen from Table 3 contains 24 basically significant items only one, namely, *rich* +24, emerges as maximally effective

It has been pointed out in a previous connection that the 92 items appearing in Table 3 are considered as basically significant in differentiating between low-high *DI*s. Twenty-two or almost one-fourth of that number have been listed as maximally effective. Certain characteristic groupings of these maximally effective items are of interest First, all nine items from Test I are definitely weighted in the direction of the least serious delinquent group. Second, the entire array of 12 items drawn from Tests II and III are weighted toward the most serious group of delinquents The single maximally effective item from Test IV has a plus denotation and, therefore, is weighted in the direction of high *DI*s.

Not unimportant is a practical consideration which presents itself with respect to the array of maximally effective items, namely, implications in the diagnosis and differentiation of delinquent conduct. One practical standpoint may be phrased thus: To what extent may maximally effective items be employed as a means of predicting probable seriousness of delinquent careers? A few elementary statistical procedures were used to provide at least a partial answer to the query.

The first course followed was to re-score *Interest-Attitude Tests* on the total group of 316 subjects in terms *exclusively* of maximally effective items from each test. Two facts are of conspicuous concern in understanding the re-scoring technique First, responses to items were indicated either by a single or double cross Second, certain maximally effective items were directionally designated by minus and others by plus signs. Hence, to re-score the tests differentially, account was taken of both foregoing points. An illustrative case will clarify the re-scoring procedure

Case M.T D., a 15-year old delinquent boy responded as follows to maximally effective items of Test I: XX *talking back* (—), XX *smoking* (—); X *pool rooms* (—), X *lawlessness* (—) No response was elicited to the items *tobacco, quarreling, fighting, playing cards,* and *shooting craps.* Summing the responses, single crosses counting one, double crosses two, the result is —6. The differential score for Test I will always be a minus quantity because the sign of all maximally effective items is minus. The highest differential score possible on Test I is —18, or the result obtained when all maximally effective items are indicated by two crosses. On Test II responses are: XX *poison* (+), XX *jail* (+), XX *cyclones* (+), XX *operation* (+). Responses were made to all maximally effective items of this test, the differential score being +8, since all items have plus denotations and are double-crossed. Following the same principle of re-scoring of Test III, a differential score of +11 resulted. The sole maximally effective item, *rich*, of Test IV was not indicated by the subject, so that the differential score for the last test was zero. Adding the four differential scores —6, 8, 11, and 0, the differential weighted score for the total test is 13. Algebraic addition was necessitated in the computation of most differential weighted scores on account of minus items in Test I. The lowest possible differential weighted score was —18 or the sum of nine double-crossed maximally effective items in Test I. The highest possible differential weighted score was 26 or the sum of thirteen double-crossed, positively denoted items occurring in Tests II, III, and IV. Differential weighted scores on the 316 subjects distributed, therefore, between the extremes indicated.[4]

A Pearson correlation between differential weighted scores and Delinquency Indexes based on the total distribution of 316 cases yielded an *r* of .34 ± .03. When the contrasted 64 least and 64 most delinquent subjects were eliminated from the total distribution, leaving approximately the middle 60 per cent or 188 cases, the coefficient for the restricted range was .16 ± .05. Accordingly, while the whole range of differential weighted scores cannot be used with unqualified assurance of predicting probable degrees of delinquency, it is not unlikely that lowest and highest differential weighted scores may provide a basis for forecasting least-most serious careers of juvenile delinquency.

Considering further the restricted distribution, mean differential weighted scores were computed for 42 subjects constituting the highest *DI*s and for 54 subjects constituting the lowest *DI*s from the distribution of 188 subjects. Delinquency Indexes of the most serious delinquents ranged approximately from 160 to 170, of the least serious delinquents from 130 to 140. The mean differential weighted score of the former group was 23.34, of the latter 20.91, giving an obtained difference of 2.43. Computing $\frac{D}{PE\ (diff)}$ the ratio was found to be 3.5, which, while not insuring complete reliability, is high enough to be considered of some significance. Conservatively interpreted it means probably that within the limits of extremely high or low differential weighted scores on the *Interest-Attitude Tests* are confined potentialities or lack of them for incurring the most or least serious degrees of juvenile delinquency. The practical conclusion in this regard would be that an individual having a differential weighted score falling within the upper class intervals of the distribution would be more likely to become a profoundly serious delinquent problem, whereas a differential weighted score occurring in the lower class intervals would indicate a tendency to a moderate degree of social offensiveness

---

[4] To avoid use of negative values a constant of 20 was added to each differential weighted score, making a possible range of 2 to 46.

A legitimate question is: What quantitative limits at the upper and lower extremes of the distribution of differential weighted scores give the most adequate basis for predicting future least-most serious careers of delinquent behavior? Any sort of answer in terms of the present analysis is conjectured from a few statistical facts. Aside from the critical ratio already stated, differential weighted scores were obtained on the two groups contrasted for degrees of seriousness in delinquency, the 64 subjects having lowest *DI*s and 64 with highest *DI*s from the total distribution of 316 subjects As previously pointed out these contrasted groups were employed for selection of basically significant and maximally effective items The mean differential weighted score of the former group was 18.66, for the latter 24.66, the difference being 6.0 found to be completely reliable since $\frac{D}{PE\ (diff)}$ equals 9.2 Interpreted broadly the last finding substantiates a previously stated conclusion, namely, that individuals having differential weighted scores of greatest magnitude are more likely to become extremely serious delinquent problems, the converse tendency holding in the case of subjects having lowest differential weighted scores. One conspicuous fact should be noted in relation to the critical ratios shown: when the critical ratio for a difference in mean differential weighted scores is computed on the basis of contrasted groups having *DI*s of 160 to 179 and 130 to 140, respectively, $\frac{\text{D}}{\text{PE (diff)}}$ is 3.5, but when the contrasted groups are constituted of subjects from class intervals 170 to 250 and 105 to 130, respectively, the critical ratio is 9.2. This would seem to mean that predictions of least-most serious delinquent careers will be accurate in proportion to the magnitude of the differential weighted score.

Inasmuch as accuracy of prediction in relation to degree of juvenile delinquency depends presumably on the magnitude of differential weighted scores on the *Interest-Attitude Tests*, a not unimportant aspect of the problem is the probability of a given individual having a certain differential weighted score becoming a more or less serious juvenile offender. To throw some light on this question an empirical criterion has been adopted in pursuance of which necessary fundamental facts are displayed in Table 4.

TABLE 4

Comparison of Least-Most Serious Delinquent Groups in Terms of Probability

Group No	Groups contrasted	N	Mean diff wtd score	SD	Value $\pm 1.29\sigma$ from mean
I	*DI*s 130-140	54	20.91	4.37	$+1.29\sigma = 26.55$
II	*DI*s 160-170	42	23.34	5.46	$-1.29\sigma = 16.30$
III	*DI*s 105-130	64	18.66	5.70	$+1.29\sigma = 26.01$
IV	*DI*s 170-250	64	24.66	5.33	$-1.29\sigma = 17.78$

A point plus or minus 1.29σ of the normal probability curve corresponds to 40.15 per cent of the surface. All distributions of differential weighted scores approximated the normal, so the application of this statistical device seems warranted. Considering Group I it was found that a distance of 1.29σ above the mean resulted in a value of 26.55. Performing the same computations for Group II, a distance -1.29σ from the mean resulted in a value of 16.30. In brief, 90 per cent of low *DI*s (Group I) fall below a differential weighted score of about 27, while 90 per cent of high *DI*s (Group II) exceed a differential weighted score of 16. Chances are 9 in 10, therefore that least serious delinquents will score less than 27, and in the same ratio most serious delinquents will score more than 16 on the maximally effective items. Groups III and IV present essentially the same picture, i.e., chances are 9 in 10 that low *DI*s will score less than 26 and in the same proportion high *DI*s will score in excess of approximately 18.

Practically the only interpretation possible from the last set of facts is that extreme differential weighted scores provide *somewhat* of a basis for predicting degree of delinquency. If the differential weighted scores are 16 to 18 or less, it is reasonable to suppose that subjects making such scores tend toward less serious degrees of offensiveness, while subjects making scores of 26 or 27 and in excess of these values tend toward more serious degrees of offensiveness. This reasoning is supported apparently by certain facts shown in Table 5.

TABLE 5

DIFFERENTIAL WEIGHTED SCORES DISTRIBUTED BY SUB-DIVISIONS BASED ON HYPOTHETICALLY PERFECT NORMAL DISTRIBUTION

Sub-division	Class interval	F Low *DI*s	% Low *DI*s	F High *DI*s	% High *DI*s	Diff	$\frac{D}{\sigma \text{ diff}}$
A	44-48	0	0	0	0	0	0.0
B	39-43	0	0	0	0	0	0.0
C	35-38	0	0	1	2	2	1.2
D	30-34	0	0	9	14	14	3.3
E	24-29	14	22	25	39	17	2.1
F	19-23	14	22	21	33	11	1.4
G	14-18	23	36	8	13	23	3.1
H	10-13	11	17	0	0	17	3.6
I	5- 9	1	2	0	0	2	1.2
J	0- 4	1	2	0	0	2	1.2

As has been pointed out the possible range of differential weighted scores is from —18 to 26, or with a constant of 20 added to eliminate negative quantities, from 2 to 46. Within this range a hypothetically perfect normal distribution, contained sixteen 3-unit class intervals, was assumed, the mean of which was 24, the standard deviation 8. Subdivisions with corresponding limiting values of differential weighted scores shown in Table 5 represent successive distances of 6 sigma each along the base line of the hypothetical distribution, assumed to extend from —3 sigma to +3 sigma. The 64 subjects from the total distribution having lowest *DI*s (105-130) and a like number having highest *DI*s (170-250) were distributed separately with reference to the subdivisions in Table 5. Employing the Edgerton-Paterson technique (5) critical ratios were computed for differences in proportion of contrasted subjects falling in each subdivision

There are too few cases in the subdivisions of Table 5 for making definite generalizations. Certain suggestive trends, however, may be observed First, a significantly larger proportion of low than high *DI*s are found in subdivisions *G* and *H* with no high *DI*s having differential weighted scores in *I* and *J* Second, a significantly greater percentage of high than low *DI*s are found in subdivisions *D* and *E* with a total absence of differential weighted scores for low *DI*s above subdivision *E*. Specifically the facts just cited appear to add a final note in support of employing differential weighted scores below and above certain values as a possible means of forecasting least-most serious degrees of juvenile delinquency. As mentioned previously, the crucial values are 16 to 18 or less for forecasting least serious and 26 to 27 or more for forecasting most serious delinquent behavior.

THEORETICAL IMPLICATIONS

Major emphasis so far has been placed on the role which traits play in quantitatively differentiating least from most serious juvenile offenders Suggestions have been made as to possible ways in which traits may be employed for predicting degrees of delinquency. An equally consequential aspect of the problem concerns the constitutional patterns in the personality structure of juvenile delinquents which differentiating traits symbolize.

The most conspicuous characteristic displayed by basically significant items related to blameworthy circumstances (Test I) are the directional tendencies of the differences between low and high *DI*s (see Table 3) Twenty-two items bear minus designations and only one a plus denotation Such preponderant weighting in the direction of least delinquent subjects suggests strongly that heightened sensitiveness to wrongs is a concomitant of moderate degrees of social offensiveness Contrariwise the most serious group of delinquents, represented by high *DI*s, are preoccupied with only one definite conception of wrong, *yellowness,* a term the intense ignominy of which is well understood by individuals who have long indulged in offenses against the social group.

Contrasted with things considered wrong, weighted singularly toward least serious delinquency, are traits referring to fears, worries, and anxieties, heavily weighted in the direction of most serious delinquency. Table 3, showing results from Test II, indicates that of 26 basically significant items, all but two have positive notations In this connection it is worthy of note that Bridges and Bridges (1) and Courthial (2) found a larger number of worries among delinquents than non-delinquents Interpreted broadly, present findings may indicate that not only are delinquents afflicted with an increased number of fears but with increase in degree of social offensiveness there develops likewise an exaggeration of fear and anxiety states

Only three, from a total of 19 basically significant items from Test III (consult Table 3), have minus denotations. Opposed to interests, such as *shooting, movie star, card parties, soldiers,* and others which are markedly differential for high *DI*s the low *DI*s are distinguished by the three interests, *dominoes, children,* and *Sunday School.* Evidently the extremely serious delinquent does not gratify his impulses through participation in cloistered pastimes. One might conclude almost that one of the chief characteristics of personality possessed by subjects with a high degree of delinquency is their attachment to interests having an inferior level of social permanency.

Basically significant items involving characteristics admired in others (Test IV) are 24 in number, weighted toward low *DI*s—fifteen being minus and nine plus (note Table 3). While the distinction is not definite as to the characteristics of the items weighted toward one extreme or the other of delinquent behavior, the group of high *DI*s seem to be more influenced by the factor of personal prestige than low *DI*s. Sample plus items are *rich, inventive, expert, wealthy, up-to-date;* sample minus items are *generous, having initiative, discreet, patient, innocent, well-informed, humorous.* The responses of least serious offenders appear to center around desirable personal traits without reference to a framework of flamboyancy.

SUMMARY

Concisely stated, the present analysis has been an attempt to discover how from four different points of view least-serious delinquents differ from most-serious delinquents. The groups contrasted for degrees of delinquency were selected by means of a Delinquency Index To accomplish the objective adopted, a technique which incorporates four separate arrays of personality traits was employed, the *Interest-Attitude Tests.* Certain definite differences were found between least and most offensive juvenile offenders as regards circumstances considered wrong, fear and anxiety states, things in which interested; and traits admired in others. In general blameworthy circumstances and qualities admired in other people are most diagnostic of least-serious delinquency, while fear or anxiety states and interests are conspicuously differential for most-serious delinquency Evidence is adduced to show that differential weighted scores based on traits maximally effective in differentiating least-most serious delinquents have a fairly significant correlation with the Delinquency Index. Other forms of analysis than correlation indicate that within certain statistical limits, differential weighted scores below or above given crucial values may possibly be used as a means of forecasting least or most serious delinquent careers. In so far as differentiating traits are cues to the fundamental patterns constituting personality, limited theoretical interpretations of the personality of least-most serious juvenile offenders has been undertaken.

REFERENCES

1. BRIDGES, J. W, & BRIDGES, K M B A psychological sudy of juvenile delinquency by group methods. *Genet Psychol. Monog,* 1926, **1**, 411-506.
2. COURTHIAL, A Emotional differences of delinquent and non-delinquent girls of normal intelligence *Arch of Psychol,* 1931, No. 133, 5-102.
3. DUREA, M A An experimental study of attitudes toward juvenile delinquency. *J Appl Psychol,* 1933, **17**, 522-534
4. ———. A quantitative method for diagnosing the seriousness of asocial behavior of juvenile delinquents *J. Gen Psychol.,* 1936, **14**, 412-421
5. EDGERTON, H A, & PATERSON, D G Table of standard errors and probable errors of percentages for varying numbers of cases. *J Appl. Psychol.,* 1926, **10**, 378-391.
6. PRESSEY, S. L., & PRESSEY, L C Development of the interest-attitude tests *J Appl Psychol,* 1933, **17**, 1-16

Department of Psychology
Ohio State University
Columbus, Ohio

The Journal of Genetic Psychology, 1938, 52, 285-302

THE CONSTANCY OF MENTAL TEST PERFORMANCE DURING THE PRESCHOOL PERIOD*

Institute of Child Welfare, University of California

MARJORIE PYLES HONZIK[1]

A problem of both practical and academic interest is the predictive value of preschool mental tests. This problem has practical implications for parents and agencies placing children for adoption, and is of academic interest in its contribution to our knowledge of the nature and regularity of mental growth in the early years.

A number of investigations have been made in this field during the past ten years. Results have been reported in the following ways: (*a*) per cent of the group whose IQ changed by specified amounts on a retest; (*b*) average change in *IQ*, and (*c*) test retest coefficient of correlation.

Terman's (16) figures show that of children first tested before the age of six years, 14 per cent showed a change of fifteen points of *IQ* or more on a retest, while of children first examined after the age of six, only 4.7 per cent varied to this extent. Johnson (12) reported that in a group of 125 children first tested between two and eight years, 18 per cent showed changes of 10 *IQ*-points or more on retests. The test retest correlation of *IQ* for this group was .80. Johnson notes that the children first tested at three years or younger show greater instability in their scores than do the older children. Hildreth (9) found changes of 20 points or more in 19 per cent of the children first tested before the age of six, but only 4.3 per cent of those first tested at a later age changed this much. Goodenough (8) reported that the test-retest correlation for three groups of children (one hundred cases in each group) aged two, three, and four years was .82 after an interval of 5.9 weeks. Fifty-six of these

---

*Accepted for publication by Harold E. Jones of the Editorial Board, and received in the Editorial Office on September 3, 1937.

[1] The writer is indebted to all who have made this study possible: the Director of the Institute of Child Welfare at the University of California, Professor H. E. Jones, for administrative assistance and many helpful suggestions; the mental testers, especially Miss Lucile Allen and Mr. G. V. Sheviakov who gave most of the tests, and the Guidance Clinic, under the direction of Dr. Jean W. Macfarlane, which has maintained the interest and cooperation of the children and parents included in the study (14).

TABLE 1

THE CONSTANCY OF MENTAL TEST PERFORMANCE OF PRESCHOOL CHILDREN

Author and date	Sample	Scale	Age in mos. on initial test	N	Interval between tests	Reported changes in mental test scores on retests: Test retest correlation	Av. diff	Range of change in scores	Percent showing specified changes in scores	Remarks
Terman (16) 1919	"From the vicinity of Stanford University"	Stanford-Binet	36-71	99	From 1 day to 7 years	r		—20 to 20	14% changed 15 *IQ* points or more	
Baldwin and Stecher (1) 1924	Iowa Preschool Laboratories	Stanford-Binet	24-84	59	3 to 20 mos				17% changed 15 per cent of the first *IQ* or more	
Johnson (12) 1925	*Private and public* schools, New York City	*Stanford-Binet*	*24-96*	125		*80*		—*17 to 32*	15% changed 10 *IQ* points or more	Children tested at 3 years or younger show greater instability in their scores than the older preschool children
Woolley (19) 1925	(a) Merrill-Palmer School (b) Merrill-Palmer waiting list	Stanford Binet	30-60	(a)43 (b)36	7 to 14 mos				(a) 33% changed 20 *IQ* points or more (b) 22% changed 20 *IQ* points or more	
Hildreth (9) 1926	Lincoln School Teachers College, Columbia	Stanford-Binet	36-60	132	"Less than a month" to 8 years			—30 to 51	19% changed 20 *IQ* points or more	
Goodenough (8) 1928	Representative Minneapolis sample	Kuhlmann Binet	24, 36, 48	300	5 9 weeks	82± 01	8.5 *IQ* points		4% changed 25 *IQ* points or more	56 of these children were retested after a year—the r's between the 1st and 3rd tests were .76 (for a younger group) and 84 (for an older group)
Updegraff (17) 1932	Iowa Preschool Laboratories	Kuhlmann or Stanford Binet	19 66	(a)63 (b)123	(a) 6 months (b) 6 months	(a) 54 (b) 84			50% changed 10 or more *IQ* points Changes were larger for children 19 to 42 months	Group (a) was first tested before entrance to the laboratory school, group (b) was given both tests after entrance
Mowrer (15) 1933	Johns Hopkins Child Institute	Stanford-Binet	20-66 (approx)	95	Av 12 7 mos		13 7 *IQ* points	—24 to 29		
Bayley (3) 1933	Superior sample of Berkeley children	California Preschool Schedule I	21, 24, 27, 30 and 36*	61	(a) 3 months (b) 6 months	(a) av 82 (b) " 79				*The same children were retested at these ages
Driscoll (5) 1933	Child Development Institute, Teachers College	Kuhlmann Binet	12-48	(a)109 (b) 91 (c) 38 (d) 23	(a) 6 months (b)12 months (c)18 months (d)24 months	(a) 61 (b) 60 (c) 56 (d) 60				Correlations between scores earned at specific ages range from 56 to 76 on the Kuhlmann-Binet and from 47 to 60 on the Merrill Palmer Scale
Kawin (13) 1934	"90% were from five Chicago Nursery Schools 10% Clinic cases"	Stanford-Binet	36-90	114(1x2)* 39 (2x3) 39 (1x3)	From 1-40 mos (1x2) 6-26 mos (2x3)	75(1x2) 84(2x3) [illegible]	9 1(1x2) 9 6(2x3)	—34 to [illegible] (1x2) —[illegible] (2x3)		*(1x2) signifies a comparison of an initial and a second test

children were given a third test a year later. The correlations between the first and third tests were .76 for a younger group (age range 24 to 30 months), and .84 for an older group (age range 39 to 50 months). These findings, and others presented in Table 1, suggest that mental test constancy is definitely dependent on the ages of the children tested.

That mental test constancy is also related to the *interval between* tests is to be expected and has been shown in a number of studies. Goodenough (8) reported an average change on a retest of 8.5 *IQ* points after an average interval of 5.9 weeks, while Mowrer (15) reported an average change in *IQ* of 13.7 when the interval between tests was a year on the average. This difference, 8.5 points of *IQ* as compared with 13.7, is probably largely dependent on the difference in the interval between tests, 5.9 weeks as compared with one year. Driscoll (5) reported test-retest correlations after intervals of 6, 12, and 24 months. A slight tendency is noted for the coefficients to decrease with an increasing interval between tests (see Table 1). It seems probable that her results would have been even more clear-cut if her findings had been reported for children of uniform age on the first test. The two factors, age at time of testing and interval between tests, were taken into consideration in a study reported by Bayley (3). The 61 children included in her group were tested at one-month and later at three-month intervals during the first three years of life. The children were all tested as nearly as possible on their monthly birthdates. Bayley obtained average test-retest correlations during the preschool period (21 to 36 months) of .84 for an interval of three months (21x24, 24x27, and 27x30 months); and a slightly lower average correlation of .83 for an interval of six months (21x27, 24x30, and 30x36).

Consideration of studies that have been made of mental test constancy during the preschool period reveals the need of a study of a fairly large group of children, of known socio-economic status, who would be tested and retested at specified ages during the preschool years. Such a study has been made on a group of children which is representative of the children living in Berkeley, California.

THE PRESENT STUDY

The Sample

The 252 children included in the present study[2] were given their first mental test at the Institute of Child Welfare at the age of 21 months At this time the group was divided into two matched subsamples of 126 children on the basis of certain socio-economic factors (parents' nationality, income, father's occupation, socio-economic rating, neighborhood, and mother's age and education). One of these subsamples was admitted to the Institute of Child Welfare's Child Guidance Clinic (under the direction of Dr. J. W Macfarlane, 14). This group will be referred to as the *"Guidance Group;"* the children in the second subsample will be referred to as the *"Control Group."* The children in the two groups were given mental tests at the following ages;

---

[2] This group of 252 children is a representative subsample with respect

Guidance	*Control*
21 months	21 months
2 years	—
2½ years	—
3 years	3 years
3½ years	3½ years
4 years	4 years
5 years	5 years
6 years	6 years
7 years	7 years

Every effort was made to test the children as nearly as possible on their birthdays. Actually from 72 to 95 per cent of the children were tested within one month of their birthdates at the various age levels. The age range included at each age level and the per cent of cases tested within a month of their birthdays are given in Table 2.

As was to be expected in a cumulative study, a number of children were unable to come in for one or more of the mental tests. The number of children actually tested at each age level is given in Table 2. There appears to be no reason to believe that the children *not* tested at any age level constitute a selected group The most frequent cause of a missed test was the family's being out of town.

The Mental Test Data

Two intelligence tests were used. Each child was tested at successive age levels on the same test, either the California Preschool Schedule I (Cal I) or the California Preschool Schedule II (Cal.

---

to certain socio-economic factors of the children included in the Berkeley Survey. The Berkeley Survey is comprised of every third child born in Berkeley over an eighteen-month period, January, 1928 to June, 1929 (18).

TABLE 2
AGE RANGE INCLUDED AT SPECIFIED LEVELS

Age level	Range of ages included (in mos)	Total number of children tested	Per cent of cases tested within one month of birthdate
1-9	20 0-22	234	95
2-0	22 5-25	113	72
2-6	25 5-33	114	77
3-0	33 5-39	229	76
3-6	39 5-45	215	80
4-0	45 5-51	211	89
5-0	57 5-66	210	81
6-0	66 5-78	214	86
7-0	78 5-91	208	95

II).[8] The number of children tested on these scales is shown in Table 3 It will be noticed that more tests were given at age level 21 months than there are children; this is because 20 children were tested on both scales at that age level

The mental test scores for each scale reported in this study are sigma scores based on the means and standard deviations of the total point scores at a given age. In some cases, for greater convenience, the sigma scores have been translated into scale scores (Scale score $= 10 \frac{x}{\sigma} + 50$, where $\frac{x}{\sigma}$ is an individual child's sigma score).

The reliability of the two preschool tests (Cal I and Cal. II) may be shown in the following ways: (*a*) correlation between scores earned by the same children on the two different tests, (*b*) split-half reliability coefficients or (*c*) test retest correlations Sixty-one children were tested on both Cal. I and Cal II between the ages of 21 months and 3 years The correlation between scores earned on the two tests was 71 Split-half reliability coefficients have not been ascertained for this group However, Bayley (2, 3) reported split-half reliability coefficients for the 60 children in the

---

[8] The published *California Preschool Scale Form A* (11) is composed largely of items from the *California Preschool Schedule I*, together with a few items from the *California Preschool Schedule II* The test items for the *California Schedules I* and *II* include selections made by Dr Adele S Jaffa from several standardized tests [Gesell (6), Minnesota (7), and Bayley (4)], together with some original items first validated at the Institute These scales have been normed by the Thurstone method of absolute scaling

Infant Growth Study on Cal. I between the ages of 21 months and 5 years. The reliability of the total test (split-half using the Spearman-Brown correction) was:

.83 at 21 months
.80 at 2 years
.89 at 2½ years
.84 at 3 years
.92 at 3½ years
.94 at 4 years
.95 at 5 years

It seems reasonable to suppose that similar split-half reliability coefficients would be obtained for the 252 children in the present study on this scale. Evidence regarding the comparative reliabilities of the two scales will be given when the constancy of test performance on the two tests is discussed.

At 6 and 7 years, all the children were tested on the Stanford-Binet. The *IQ*'s obtained were converted into sigma and scale scores so that they would be in comparable form to the mental test scores obtained between 21 months and 5 years.

The children were tested by experienced examiners, and usually by the same examiner on successive tests. Miss Lucile Allen tested the majority of children in the Control Group at every age level; Mr. Sheviakov and Mrs. Campbell tested the Guidance children whom they were following in the Guidance Clinic.

Constancy of Mental Test Performance.

The constancy of mental test performance on the two preschool mental tests, the *California Preschool Schedules I* and *II*, is presented in Table 3. Correlations between children's scores at adjacent age levels are fairly high, varying from .64±.05 to .76±.03, but the correlations decrease with the interval between tests. There is only very slight prediction of performance on the Stanford-Binet at 6 or 7 years from an initial test given at 21 months (*r*'s from .17 to .32)—and *this is true for both tests* (Cal. I and II). However, the 3-year, 3½-year, 4-year and 5-year test scores do give a fair prediction of performance on the Stanford-Binet at 6 and 7 years (*r*'s range from .54±.05 to .76±.03).

The higher intercorrelations of Cal. I scores at the younger ages suggest that this test is a more reliable one between the ages of 21 months and 4 years. However, Cal. II scores at 5 years are more

TABLE 3

Comparison of the Consistency of Mental Test Performance of California Preschool Schedules I and II

									STANFORD-BINET	
AGE			2-0	2-6	3-0	3-6	4-0	5-0	6-0	7-0
	Test	*n*	*r*	*r*	*r*	*r*	*r*	*r*	*r*	*r*
1-9	Cal I	181	74± 03	65± 04	58± 04	59± 04	41± 05	39± 06	32± 05	30± 05
	Cal II	73			46± 06	35± 07	34± 07	42± 07	17± 08	26± 08
3-0	Cal I	153				77± 02	64± 04	60± 04	57± 04	57± 04
	Cal II	94				64± 05	46±.06	54± 06	56± 05	54± 05
3-6	Cal I	132					82± 02	73± 03	63± 04	59± 04
	Cal II	94					69± 04	68± 04	71± 04	58± 05
4-0	Cal I	121						71± 03	62± 04	59± 04
	Cal II	102						71± 04	61± 04	58± 05
5-0	Cal I	108							67± 04	69± 04
	Cal II	104							76± 03	.76± 03

highly correlated with scores on the Stanford-Binet at 6 and 7 years than are the Cal. I scores. Thus it would seem that Cal. I is a more reliable test and that it has a greater predictive value at the earlier ages, while Cal. II is a more reliable test at the later ages of the preschool period. There is no evidence that the difference in constancy figures for the two tests is due to differences in variability.

Probably the most interesting fact to be gleaned from Table 3 is the similarity in the trend of the results for the two tests. These two entirely different tests (although similar in form and type of test items) agree in showing that constancy of mental test performance is markedly dependent on both the ages of the children when tested and on the interval between tests

Age Ratio

One way in which the dependence of test constancy on interval between tests and age of testing may be shown is to consider the intercorrelations between mental test scores in relation to the age ratio,

$$\frac{\textit{CA at first test}}{\textit{CA at second test}}$$

It is clear that the age ratio decreases as the interval between tests increases. Thus,

$$\frac{\textit{CA of 1 year at first test}}{\textit{CA of 2 years at second test}} = 50 \text{ is greater than } \frac{\textit{CA 1 year at first test}}{\textit{CA 3 years at second test}} = .33.$$

But with the *interval between tests held constant,* an age ratio at a higher age level is greater than an age ratio at a lower age level. Thus,

$$\frac{\textit{9 years}}{\textit{10 years}} = .90 \text{ is greater than } \frac{\textit{2 years}}{\textit{3 years}} = .66$$

The magnitude of a correlation between tests varies directly with the age ratio, that is, the higher the ratio the greater the correlation This is shown in Table 4. The correlation between 22 age ratios and the corresponding correlation coefficients (for Cal I) is .92± 02. Thus a high correlation between test scores may be due to a short interval between tests at a lower age level or to a higher age of testing with a longer interval between tests. With the same interval between tests, the correlation will be higher at the higher age levels. It should be emphasized that the relation found here between the age ratio and mental test intercorrelations was for the age range 21 months and 7 years. It is quite possible that this

TABLE 4

SCATTER DIAGRAM SHOWING THE RELATIONSHIP BETWEEN THE AGE RATIO AND CAL. I INTERCORRELATIONS OF MENTAL TEST SCORES

	AGE RATIO*													TOTAL
	25—29	30—34	35—39	40—44	.45—49	50—54	.55—59	60—64	65—.69	70—.74	75—79	.80—84	85—89	f
80—84													1	1
75—79													1	1
70—74										1		1	1	3
65—69										2		1		3
60—64							1	1	1		1			4
55—59				1		3	2							6
50—54														—
45—49														—
40—44				1										1
35—39			1											1
30—34	2													2
Total f	2		1	2		3	3	1	1	3	1	2	3	22

$*\frac{\text{CA at First Test}}{\text{CA at Second Test}}$

$r = 92 \pm 02$

degree of relationship would not be found to exist for the entire growth period from birth to maturity. However, if the relationship is high throughout the growth period, the age ratio does afford a simple means of recalling the magnitude of the relationship which exists between mental test scores given at different age levels The correlation between the age ratio and the correlations obtained for Cal II is 78±.06, which is considerably lower than that found for Cal I. It is possible that this much lower coefficient invalidates the age ratio concept, but it is also possible that the relationship is lower because the Cal II scores are less reliable. Inspection of the coefficients obtained for Cal. II shows more inconsistencies than for Cal I The *r* having the most effect in lowering the relationship between the age ratio and the *r*'s is the correlation between the 3- and 4-year test scores of .46 (see Table 3). Higher coefficients were obtained between the 3-year scores and scores obtained at 5, 6, or 7 years, suggesting that a higher coefficient was to be expected between the 3-year and 4-year test scores.

In Table 5, the constancy of mental test performance for the Guidance and Control Groups is given, regardless of whether children were tested on Cal. I or II. The trends noted when considering results for Cal. I and Cal II are to be seen again when results for these two different groups of children are compared. The average of the intercorrelations for the Guidance Group is 57 (excluding results obtained at 2 and 2½ years), and for the Control Group .59 Correlations between scores earned at adjacent age levels are fairly high, varying from .65 to .77. The tendency is again noted for the correlations to vary directly with the ages of the children and inversely with the intervals between tests Thus the correlations for the Guidance Group over the following two-year intervals are:

2 to 4 years = *.46*,
3 to 5 years = *53*;
4 to 6 years = *.61*; and
5 to 7 years = *73*

There are several slight inconsistencies in the results obtained for the two groups. A higher correlation was found between the 21-month and 3-year tests for the Control Group (.59), than for the Guidance Group (47). The reverse might have been expected since a larger proportion of the Control children were tested on the less reliable Cal. II. However, it seems that the coefficient for

TABLE 5

THE CONSTANCY OF MENTAL TEST PERFORMANCE IN THE GUIDANCE AND CONTROL GROUPS

AGE			2-0	2-6	3-0	3-6	4-0	5-0	STANFORD-BINET 6-0	7-0
	Group	*n*	*r*	*r*	*r*	*r*	*r*	*r*	*r*	*r*
1-9	Guidance	117	68± 04	.59± 04	47± 05	50± 05	46± 05	32± 06	30± 06	42± 06
	Control	117			59± 04	.47± 05	33± 06	.43± 06	30± 06	19± 07
2-0	Guidance	113		71± 03	69± 03	60± 04	46± 05	32± 06	47± 05	46± 05
2-6	Guidance	114			73± 03	64± 04	57± 05	46± 05	37± 06	38± 06
3-0	Guidance	116				.73± 03	64± 04	53± 05	54± 05	56± 05
	Control	113				71± 03	59± 04	60± 04	60± 04	54±.05
3-6	Guidance	107					78± 03	.72± 03	59± 04	63± 04
	Control	108					.74± 03	.71± 03	62± 04	.59± 04
4-0	Guidance	105						69± 04	61± 04	66± 04
	Control	106						75± 03	62± 04	.53± 05
5-0	Guidance	104							65±.04	73± 03
	Control	106							77± 03	72± 03
S B 6-0	Guidance	109								81± 02
	Control	105								83± 02
S B 7-0	Guidance	104								
	Control	104								

the Guidance Group is too low partly because of chance factors, since a higher correlation was obtained between the 21-month and 3½-year test than between the 21-month and 3-year test. Another inconsistency occurs in the correlation between the 21-month scores and those obtained at 7 years (.42). The comparatively large standard deviation of the 7-year test scores may account in part for this coefficient.

The correlation of .30 between the 21-month and 6-year mental test scores indicates the slight prediction that exists over that age period. Updegraff's (17) and Goodenough's (8) finding that an initial test is less reliable than a subsequent test should be recalled when considering these results. However, the same trend is noticeable when performance at 2 and 2½ years (which are not "first tests") are compared with those at 4, 5, 6, and 7 years.

The correlation between the age ratio and intercorrelations of mental test scores for Guidance children (36 r's shown in Table 4) is .87±.03. The correlation between the age ratio and correlations obtained for the Control Group is .91±.02. These correlations between the age ratio and constancy coefficient for the two tests (Cal. I and II) and for the two groups (Guidance and Con-

TABLE 6

RELATION OF CHANGES IN MENTAL TEST SCORES TO THE INTERVAL BETWEEN TESTS

Change in sigma score	Interval between tests 1 Year 7-6 yr.		2 Years 7-5 yr		3 Years 7-4 yr.		4 Years 7-3 yr.		5 Yrs., 3 Mos. 7 yr.-21 mo	
	n	%	n	%	n	%	n	%	n	%
3.5			1	*1*					1	*1*
3									2	*1*
2.5							1	*0*	5	*3*
2					1	*1*	6	*3*	8	*4*
1.5	3	*1*	6	*3*	17	*9*	18	*9*	12	*6*
1.	16	*8*	28	*14*	23	*12*	24	*12*	16	*8*
.5	48	*24*	39	*20*	38	*20*	35	17	33	*17*
0.0 (—.2 to .2)	62	*31*	64	*32*	41	*21*	41	*20*	40	*20*
—.5	52	*26*	34	*17*	32	*17*	38	*19*	27	*14*
—1.0	18	*9*	22	*11*	25	*13*	29	*14*	27	*14*
—1.5	2	*1*	4	*2*	11	*6*	8	*4*	10	*5*
—2.0					5	*3*	1	*0*	10	*5*
—2.5	1	*0*							3	*2*
—3.0									2	*1*
Total	202	*100*	198	*100*	193	*102*	201	*98*	196	*101*
Av. change in sigma score	.47			.56		.71	.72			.92

trol) agree fairly well in showing that the age ratio may prove to be a useful concept in relation to group mental test constancy figures.

An interesting consideration is whether the correlations between mental test scores are determined by a relatively small proportion of cases or by the group as a whole. Distributions of the changes in mental test scores between certain age levels are given in Table 6. It may be seen in this table that the tendency to change in score is a general one. The percent of cases showing a very slight change on a retest varies from 20 per cent between 21-months and 7 years to 31 per cent between 6 and 7 years. The average change in score between 6 and 7 years is *.5* of a sigma,[5] while the average change in score between 21 months and 7 years is *.9* of a sigma score. There are a few instances of marked changes in score but the distributions of changes in scores are normal and suggest a multiplicity of factors determining the changes in mental test performance.

Examples of varying degrees of constancy of mental test performance for individual children are shown in Figure 1. Case 355 and Case 385 (who earned the highest *IQ* in the group at 6 years) represent instances of extremely marked gains in relation to the group scores between 21 months and 6 years. Case 355 increased from the 1st to the 90th percentile (assuming that the distribution of sigma scores is normal at the various age levels), Case 385 increased from the 31st to the 99.998th percentile. Case 324 in contrast consistently earned one of the highest scores in the group. It is interesting to note how comparatively little her scores vary on the eight different tests. Case 405 is a microcephalic child born to a superior family. She was diagnosed a microcephalic at the age of 21 months. Her scores have been the lowest in the group at every age level.[6] Case 351 represents a fluctuating performance. Other cases could be presented to show a drop in test score followed by a gain. Case 539 score above the average at 21 months but obtained the second lowest score in the group at 6 years. Her

---

[5] The standard deviation of the 6- and 7-year *IQs* is 13 points of *IQ*. Thus an average change in sigma score of .5 may be considered to equal 6.5 points of *IQ* approximately.

[6] Case 405's scores have not been included in any of the correlations reported in this study. It is not to be expected that such a markedly deviating case would be found in a sample of 252 cases. If her scores had been included in the correlations it would have been necessary to report the coefficient "including" and "not including" her scores. The inclusion of her scores would, of course, have raised the correlations reported between mental test scores earned at different age levels.

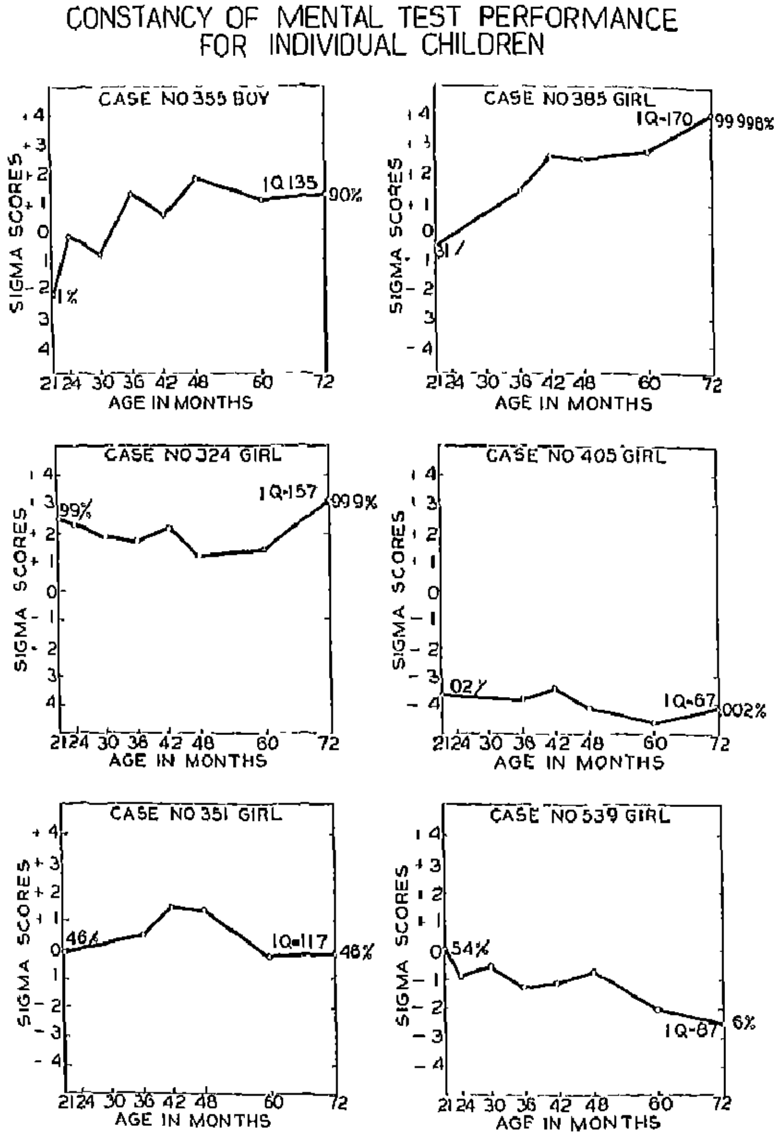

CONSTANCY OF MENTAL TEST PERFORMANCE
FOR INDIVIDUAL CHILDREN
CASE NO 355 BOY
IQ 135
90%
1%
CASE NO 385 GIRL
IQ-170
99 998%
31%
CASE NO 324 GIRL
99%
IQ-157
99 9%
CASE NO 405 GIRL
02%
IQ-67
002%
CASE NO 351 GIRL
46%
IQ-117
46%
CASE NO 539 GIRL
54%
IQ-87
6%
SIGMA SCORES
+4
+3
+2
+1
0
-1
-2
-3
-4
21 24 30 36 42 48 60 72
AGE IN MONTHS

FIGURE 1

steady decline in mental test performance with reference to the group is disturbingly impressive.[7]

The constancy of mental test performance for children earning extremely high and extremely low scores at 21 months and at 6 years have been considered in some detail. Of the ten children who obtained the lowest scores at 21 months, only two were among the ten who obtained the lowest scores at 6 years (the microcephalic Case 405 was, of course, one of these and she was the only case of the ten who failed to show some gain between 21 months and 6 years with reference to the group scores). The average sigma score of these 10 children increased from —2 4 sigma at 21 months to — 9 sigma at 6 years The greatest improvement for this group of 10 children occurred between 21 and 24 months Factors which may be contributing to the marked changes in scores between the first two tests are (*a*) unreliability of the initial test, due to emotional or other factors, (*b*) changes in the nature of the test items, that is, a larger proportion of test items requiring the comprehension and use of language at 24 than at 21 months The greatest gain in this group of children who earned such low scores at 21 months, was shown by Case 355 (Figure 1) with an increase of 3.5 sigma The average *IQ* of the 10 children earning the lowest scores at 6 years was 92 The scores of these children decreased rather gradually from an average sigma score of —1 at 21 months to an *average sigma score of —2 1 at 6 years The scores of two of the* 10 children decreased over 2 sigma.

Two of the 10 children earning the highest scores at 21 months were among the 10 highest at 6 years Thus, for both the low and the high group, two out of 10 (or 20 per cent of the children) maintained their extreme positions during the period under consideration Case 324 (Figure 1) is one of those who maintained her extremely high scores. The average scores of these children dropped from 2.2 sigma to 1. sigma, and as in the case of the children with the lowest scores, the greatest change took place between 21 and 24 months. Two of the 10 children showed a decrease of 2.8 in sigma score.

---

[7] It would be appropriate at this point to attempt some explanation of the facts that have been presented However, since many of the possible explanations (emotional factors, environment stimulation of the home and nursery school, etc) can be evaluated from the available data, this discussion of determining factors will be postponed to a future publication in which the relation of certain factors to mental test scores and changes in mental test scores will be presented (cf 10)

The 10 children with the highest scores at 6 years increased from an average of .6 sigma to 2 4 sigma between 21 months and 6 years. This increase took place quite gradually from one age level to the next. Case 385 (Figure 1) showed the greatest gain (4.6 sigma).

Discussion of the constancy of mental test performance for children earning extreme scores shows that marked changes occur over this age period, and indicates further that the changes are most apt to occur between 21 and 24 months In other words, the initial test does seem to lack the predictive value of a second test given fairly soon after the first.

The results reported on the constancy of mental test performance for children earning extreme scores show that there is a strong tendency for a regression toward the mean on retests The correlation between high scores on the first test at 21 months and gains in mental test scores between 21 months and 6 years is —.68 Thus, for a child earning an extreme score at 21 months, there is more likelihood of a regression toward the mean than constancy of performance in the subsequent years.

SUMMARY

A group of 252 children who comprise a representative sample of the children living in Berkeley, California, were given mental tests at specified ages during their preschool years. The children were first tested at the age of 21 months, and subsequently on their half-yearly and yearly birthdates. The constancy of mental test performance between 21 months and 7 years was found to be essentially the same on two different preschool mental tests (*California Preschool Schedules I* and *II*), and for two comparable subsamples of the main group Although test performance is fairly constant over short age periods, mental test constancy appears to be markedly dependent on both the ages of the children and the interval between tests. The age ratio,

$$\frac{CA \text{ at first test}}{CA \text{ at second test}}$$

which takes both these factors into account, is correlated with the mental test constancy coefficients on the *California Preschool*

Schedule I to the extent of 92± 02, and with test constancy coefficients on the *California Preschool Scheule II* .78±.06. The results show that an initial test given at the age of 21 months gives a negligible prediction of test success on the Stanford-Binet at 6 or 7 years. Tests given at three years or later to children who have had test experience are much more predictive of success on the Stanford at 6 or 7 years (correlations range from .58 to .76), yet are not high enough to warrant full confidence if one were adopting a child. Marked individual differences in test constancy are noted Twenty per cent of the children earning extremely high or extremely low scores at 21 months maintain that position to the 6-year test On the other hand, there are many instances of extremely marked changes in mental test scores One child gained as much as 4 sigma, while two children gained over 3 sigma between 21 months and 6 years

These results suggest the impossibility of making an accurate prognosis of future ability on the basis of a single mental test given before the age of two, but suggest further that repeated tests and tests at later ages of the preschool span have increasing predictive value

REFERENCES

1 BALDWIN, B T, & STECHER, L I Additional data from consecutive Stanford-Binet tests *J. Educ Psychol*, 1922, **13**, 556-560

2 BAYLEY, N., & JONES, H. E. Environmental correlates of mental and motor development a cumulative study from infancy to six years. *Child Devel.* (In press)

3 BAYLEY, N Mental growth during the first three years, a developmental study of sixty-one children by repeated tests. *Genet Psychol Monog*, 1933, **14**, 1-85

4. ——— Performance tests for three-, four-, and five-year-old children *J Genet. Psychol*, 1926, **33**, 435-454

5 DRISCOLL, G P The development status of the preschool child as a prognosis of future development *Child Devel Monog*, 1933, No 13.

6 GESELL, A The Mental Growth of the Pre-school Child. New York Macmillan, 1925 Pp 447

7 GOODENOUGH, F. L., FOSTER, J. C, & VAN WAGENEN, M J Minnesota preschool scale (Manual Form A) Philadelphia: Educ Test Bur, Inc, 1932

8 GOODENOUGH, F. L The Kuhlmann-Binet tests for children of preschool age Univ Minn, Inst Child Welf Monog Series, 1928, No 2. Pp. 146

9 HILDRETH, G Stanford-Binet retests of 441 school children *J Genet Psychol.*, 1926, **33**, 365-386

10. HONZIK, M. P. Factors related to changes in mental test performance during the preschool period. Ph. D. dissertation, Univ. Calif. Library, 1936.

11. JAFFA, A. S. The California preschool mental scale (Form A). Univ. Calif. Syllabus Series, 1934, No. 251. Pp. 66.

12. JOHNSON, B. Mental Growth of Children in Relation to the Rate of Growth in Bodily Development. New York: Dutton, 1925. Pp. xix-160. (See pp. 79-90.)

13. KAWIN, E. Children of Preschool Age. Chicago: Univ. Chicago Press, 1934.

14. MACFARLANE, J. W. Studies in Child Guidance. I. Methods of data collection and organization. MS at Univ. Calif. Instit. Child Welfare.

15. MOWRER, W. M. C. Intelligence scales for preschool children. *Child Devel.*, 1933, 4, 46.

16. TERMAN, L. M. The Intelligence of School Children. Boston: Houghton, 1919. Pp. 317.

17. UPDEGRAFF, R. The determination of a reliable intelligence quotient for the young child. *J. Genet. Psychol.*, 1932, **41**, 152-165.

18. WELCH, F. M. The Berkeley Survey. A study of the socio-economic status of four hundred Berkeley families in years 1928-1929. MS at the Univ. Calif. Instit. of Child Welfare.

19. WOOLEY, H. T. The validity of standards of mental measurement in young childhood. *School & Soc.*, 1925, 21, 476-482.

Institute of Child Welfare
University of California
Berkeley, California

The Journal of Genetic Psychology, 1938, **52**, 303-325

STUDIES IN MENTAL DEVELOPMENT. I. PERFORMANCE ON GESELL ITEMS AT SIX MONTHS AND ITS PREDICTIVE VALUE FOR PERFORMANCE ON MENTAL TESTS AT TWO AND THREE YEARS*

Samuel S. Fels Research Institute, Antioch College

VIRGINIA LAFAYETTE NELSON AND T. W. RICHARDS

One phase of the psychological testing program in the Fels study of child development has dealt with the use of mental tests at intervals during postnatal life. The tests used and frequency of observation have been the following:

> Gesell developmental items as described by Gesell (5) at six, twelve, eighteen, and twenty-four months
>
> Merrill-Palmer Scale at eighteen, twenty-four, thirty, thirty-six, forty-two months (9)
>
> Stanford Revision (1916) of the Binet-Simon Scale at thirty-six, forty-eight, and every year thereafter (10)

The group tested has included about 150 infants, most of whom are not yet old enough to have taken the full battery of tests An attempt was made to test the child as near as possible to his actual birthday

It is the purpose of this study, among other things, to determine the relationship between performance on early tests and that demonstrated on later tests The present report is concerned with the interrelationships between the performance on the Gesell items at six months, and the relation of these performances to the total test, the Merrill-Palmer test at 24 months, and the Stanford-Binet at 36 months. Correlation has also been determined between the total test at six months and each of these later tests.

One hundred and twenty-three children were examined at or near the six-month birthday, by four examiners. Seventy-five per cent of the examinations were made by one examiner (the senior author). Of the total group, 82 of these infants were examined within four

---

*Received in the Editorial Office on September 3, 1937

TABLE 1

PERCENTAGE OF SIX MONTHS OLD INFANTS PASSING GESELL ITEMS, COMPARED WITH PERCENTAGES OBTAINED BY OTHER INVESTIGATORS

	Fels			Percentages			
Item	No cases	Per cent	P E	Bayley	Fillmore	Linfert-Hierholzer	Gesell
Holds head erect	82	100	0.0	98			85-100
Hand reacts to table	82	100	0.0	98			85-100
Expresses recognition	82	99	0.7				65- 84
Regards cube	78	99	0.8	100			85-100
Shows conscious strangers	82	98	1.1	65			65- 84
Plays with objects	81	98	1.1	98			85-100
Dangling ring: closes in	82	96	1.5	97			85-100
Turns head to bell	81	95	1.6	98			65- 84
Crumples paper	81	95	1.6				65- 84
Blinks at pencil	77	95	1.6				65- 84
Frolics when played with	78	94	1.8				50- 64
Manipulates spoon, cup, saucer	80	91	2.2				65- 84
Sits with slight support	82	89	2.4	91			50- 64
Dangling ring clasps ring	82	89	2.4				65- 84
Holds two cubes	80	88	2.5	67			65- 84
Reacts mirror image	80	88	2.5	49	69		20- 49
Conscious fallen object	79	84	2.8	50			65- 84
Picks up cube	81	77	3.2	26			65- 84
Rolls back to stomach	80	76	3.2	34		70	65- 84
Lifts inverted cup	80	76	3.2	60	74		65- 84
Splashes in tub	74	74	3.4				65- 84
Pats table	81	74	3.3			56	65- 84

TABLE 1 (*continued*)

Item	Fels No cases	Fels Per cent	P E	Percentages Bayley	Percentages Fillmore	Percentages Linfert-Hierholzer	Percentages Gesell
Inhibits head one hand	75	65	3 7				20- 49
Reaches directly for spoon	82	65	3 6				65- 84
Manipulates one hand	73	63	3 8	37			20- 49
Makes stepping movements	75	63	3 8			6	20- 49
Looks for fallen object	78	56	3 8	14	64	48	20- 49
Throws objects to floor	75	55	3 9				20- 49
Takes bottle in and out mouth	75	52	3 9				20- 49
Music. stops crying	70	51	4 0				65- 84
Exploratory manipulations	79	51	3 8				20- 49
Dangling ring above head	82	50	3.8				50- 64
Bangs spoon	78	49	3 8	49			65- 84
Regards pellet	72	46	4 0	75		39	50- 64
Dangling ring persistent reach	82	43	3 7	45			20- 49
Drops 1 cube for 3rd	80	36	3.6				50- 64
Secures cube (inv cup)	80	31	3.5				0- 19
Music laughs	68	29	3 7				50- 64
Paper purposeful reaction	81	25	3 3				20- 49
Casts objects for noise	74	22	3 3				20- 49
Creeps or hitches	74	20	3 1	4		19	0- 19
Says "mama" or "dada"	76	18	3 0	0		25	0- 19
Prefers 1 hand	74	18	3 0				0- 19
Sits alone	80	16	2 8	4	58	36	0- 19
Pellet. whole hand	70	14	2 8	16	4		20- 49
Music coos	67	9	2 4				20- 49
Imitates drop in cup	75	8	2 2				0- 19
Fine prehension	70	4	1 6	0			0- 19

days of the birthday; of these 82, five tests were rejected because so few items received a score.

There are, at the six months level of the Gesell Series, 48 items or possible scores. The number of subjects tested for each item, the percentage passing each item, and the percentages obtained by four other investigations,—the Gesell (5), Bayley (1),[1] Fillmore (4), and Linfert and Hierholzer (7) studies,—are presented in Table 1.

It will be seen in Table 1 that *holds head erect,* and *hands react to table* were passed by 100 per cent of the Fels group. Ten other items were passed by at least 90 per cent. The most difficult items, passed by less than 10 per cent of the group, were *coos to music* (9 per cent), *imitates drop in cup* (8 per cent), and *fine prehension* (4 per cent).

An examination of the Fels percentages passing in Table 1 will indicate that there is not a uniform distribution of items in terms of difficulty throughout the scale. In terms of deciles the frequencies are presented in Table 2.

TABLE 2

Per cent	Number of items
*91-100	12
81- 90	5
71- 80	5
61- 70	4
51- 60	5
41- 50	4
31- 40	2
21- 30	3
11- 20	5
1- 10	3

*Includes two items passed by 100% of group.

There is thus an overabundance of easy tests, while the frequency of tests throughout the deciles lower than the highest is somewhat equivalent.

In order to determine the extent to which the results on our group might compare with those obtained by other investigators, correlations between the percentages of the various groups passing items apparently in common were made. Since there were so few items in common between investigations (and since the Gesell percentages are given in intervals of unequal lengths) coefficients were obtained by the method of rank differences (Table 3) They were as follows·

---

[1] Dr. Bayley does not present the percentages passing single items in her published reports (1, 2, 3) but very kindly forwarded them to us in personal communication. We wish to acknowledge her kindness.

TABLE 3

	No items	Coefficient
Percentages pass, Fels and Gesell groups	48	.85 ±.03
Percentages pass, Fels and Fillmore groups	5	.91 ± 05
Percentages pass, Fels and Linfert-Hierholzer	8	57 ± 17
Percentages pass, Fels and Bayley	24	* 588± 09
Fels percentage pass and Bayley age values	24	*— 865±.03
Fels percentage pass and Fillmore age values	5	— 91 ± 05

*Pearson coefficient.

Factors to be taken into account in comparing the Fels results with those obtained by other investigators are (*a*) that it is extremely difficult to ascertain, simply by examination of reports in the literature, whether or not the technique used for a given item, like *sits with slight support* is comparable with that used by us or by other investigators An attempt has been made to avoid comparisons where the similarity of Fels technique to reported technique was at all doubtful. (*b*) The numbers of cases used in studies compared with our own were approximately as follows: Gesell, 50 infants, Linfert and Hierholzer 46-50; Fillmore 87-139, Bayley 61. Each of the groups, with the possible exception of Gesell's, was possibly of slightly superior socio-economic status, as was the Fels group.

Statistical analysis showed that the Fels group had significantly less difficulty than the Bayley group on 12 of 24 comparable items, and significantly greater difficulty on one, *regards pellet* By comparison with the Fillmore group, one item was significantly easier for the Fels group *pellet whole hand,* and one more difficult *sits alone* The Linfert and Hierholzer group was superior to the Fels groups on *pats table,* and inferior on *makes stepping movements* Such differences may be due to real differences between the groups in ability, the observations of other investigators differing as much from each other as from the Fels observations. It would be unsafe to assume, however, that differences in ability account entirely for the variation in recorded performance, for there are undoubtedly differences in the criteria of success used in various laboratories.

The mean percentages passing comparable items in each of the investigations compared with the Fels investigation are presented in Table 4.

TABLE 4

Investigator	Number of comparable items	Fels mean per cent	Other mean per cent
Fillmore	5	50.2	53.8
Linfert-Hierholzer	8	46.3	37.4
Bayley	24	66.9	55.8
Gesell	48	61.1	54.4

The Fels group would seem, from the above, to pass comparable items slightly more easily than any of the other groups except the Fillmore group. This may be due to an actual superiority in performance, or to a lower criterion of performance on the part of the examiners for the Fels group. It is obvious from these data that the distribution of items in terms of difficulty is neither a normal nor an even one; there are too many easy tests.

SCORING THE TOTAL SCHEDULE AT SIX MONTHS

Attempts to obtain for the total test a score weighted for difficulty of the items were made by assigning probable error values to the percentages passing. When total scores thus weighted were correlated with the ratio of number of items passed to the number attempted a coefficient (Pearson) of .956±.01 was obtained, indicating that the number of items passed was as good an index of performance on the total test as was such a score weighted for difficulty.

The distribution of number of items passed (in all cases calculated

TABLE 5

Score	Frequency
6.0 to 10.9	3
11.0 to 15.9	4
16.0 to 20.9	10
21.0 to 25.9	15
26.0 to 30.9	9
31.0 to 35.9	20
36.0 to 40.9	14
41.0 to 45.9	2
Total	77
Mean	28.17
Sigma	8.70

as the number out of a possible 46) is given in Table 5. It will be seen that the distribution is definitely bimodal.

The coefficient of reliability for this group of 46 items was obtained by the split-half method, using items in two series alternating in difficulty The value obtained with 65 cases, +.79±.03 is raised by the Spearman-Brown correction to +.883±.02 This compares favorably with values reported in other investigations at six months: Linfert and Hierholzer +.81, Bayley +.95, Fillmore +.90.

The Relationship between Total Test Performance and Age

Although the group used in calculating the percentages passing each item was restricted to those tested within four days of the sixth-month birthday, there were 110 children tested in all. The spread in age is given in Table 6

TABLE 6

Age interval	Number cases
22.5 to 31.4 days early	1
13.5 to 22.4 days early	0
4.5 to 13.4 days early	4
4.4 days early to 4.5 days late	77
4.6 to 13.5 days late	17
13.6 to 22.5 days late	4
22.6 to 31.5 days late	3
31.6 to 40.5 days late	2
40.6 to 49.5 days late	0
49.6 to 58.5 days late	0
58.6 to 67.5 days late	1
Total	109

The correlations between age and ratio of passed to attempted items[2] for all children who tried at least two-thirds of the items was +.10±.06, indicating little relationship

Sex Differences in Performance on the Six Months Gesell Schedule

No reliable difference between sexes was apparent in total test scores. Statistical treatment of differences on single items yielded no critical ratios (on a sigma basis) of three, and but one greater than two (2.7, the females being superior to males on *clasps dangling ring*)

---

[2] Referred to hereafter as total test score

RELATIONSHIP BETWEEN PERFORMANCE ON SINGLE ITEMS AND PERFORMANCE ON THE WHOLE TEST AT SIX MONTHS

Biserial coefficients of correlation were determined for the group tested within four days of the sixth-month birthday between the performance on certain items and the total test score. These coefficients, for all items passed by not more than 90 per cent of the group, are presented in Table 7, together with the percentages of the group passing each item.

TABLE 7

	Biserial r*	Percentage
Inhibits head and one hand	.88	65
Exploratory manipulations	.87	51
Lifts inverted cup	87	76
Reaches directly for spoon	86	65
Secures cube (from inverted cup)	.82	31
Picks up cube	.82	77
Dangling ring: seizes above head	80	50
Dangling ring: persistent reaching	.80	43
Pats table	.80	74
Reacts to mirror image	77	88
Looks for fallen object	77	56
Throws objects to floor	75	55
Manipulates one hand	72	63
Makes stepping movements	69	63
Paper purposeful reaction	68	25
Drops one cube for third	68	36
Sits with slight support	.66	89
Dangling ring: clasps ring	.66	89
Takes bottle in and out of mouth	65	52
Bangs spoon	63	49
Regards pellet	62	46
Creeps or hitches	61	20
Holds two cubes	58	88
Prefers one hand	58	18
Music: laughs	58	29
Conscious of fallen object	.55	84
Rolls back to stomach	55	76
Casts objects for noise	55	22
Pellet whole hand	54	14
Sits alone	52	16
Says "mama" or "dada"	51	18
Splashes in tub	.47	74
Music: stops crying	.37	51

*These coefficients are about .06 points higher than those obtained by correlating the items with the total test less the item in question, the difference varying from .04 to 07 as the frequency in one category becomes more or less than 50%

Examination of the coefficients above shows that those items correlating best with the total number of items passed are of distinctly

TABLE 8

Motor	
Inhibits head and one hand	88
Exploratory manipulations	87
Lifts inverted cup	87
*Reach for spoon	86
Secure cube	82
Pick up cube	82
Dangling ring seize	.80
*Dangling ring reach	80
Pat table	80
Throw objects to floor	75
Manipulates one hand	.72
Drop one cube for third	68
Paper purposeful reaction	68
Dangling ring clasps	66
Bottle in and out mouth	65
Bang spoon	63
Hold two cubes	.58
Prefer one hand	59
*Casts objects for noise	55
Pellet· whole hand	54
Splash in tub	47
MEAN MOTOR (N = 21)	71
Awareness or Distance Perceptive	
Reaches for dangling ring	80
Reaches directly for spoon	86
Reacts to mirror image	77
Looks for fallen object	77
Regards pellet	62
Music laughs	.58
Conscious of fallen objects	55
Casts objects for noise	55
Music stops crying	37
MEAN AWARENESS OR D P (N = 7)	65
Locomotor-postural	
Makes stepping movements	69
Sits with slight support	.66
Creeps or hitches	61
Rolls back to stomach	.55
Sits alone	52
MEAN LOCOMOTOR-POSTURAL (N = 5)	61
Language	
Says "mama" or "dada"	51

*Also included in awareness or distance series perceptive more grasping and manipulative nature than are those correlating least. It is obvious that there are more items of this sort in the series at this level, which accounts for the nature of these relationships. When items are classified roughly according to five categories—motor (including reaching, grasping, manipulating, laterality), awareness or distance-perceptive, locomotor-postural, language and adaptive, the results in averaging coefficients were obtained as shown in Table 8.

Such a summary, qualitative as it is, shows definitely that there is a premium placed upon items involving grasping, reaching, and manipulative behavior, with such features as posture, locomotion, and perception, in a secondary position. This is done simply by including an overabundance of items of this motor character, and this is due possibly to the fact that such items of behavior are most characteristic or obvious at this age. A more careful statistical analysis of interrelationships between items is in order, to determine what specific factors are tested by this repertoire of behavior at this age level. Before making this analysis, however, it will be interesting to determine how the total test score on six-months Gesell items and how the items themselves are related to mental development in subsequent years.

THE RELATIONSHIP BETWEEN PERFORMANCE ON THE GESELL SERIES AT SIX MONTHS AND PERFORMANCE OF TESTS AT LATER AGE LEVELS

Longitudinal correlations between early test performance and later *IQ* have been obtained by Muhlenbein (8) in connection with the Linfert-Hierholzer material, by Bayley (1) for her series, and by Fillmore (4), who used the earliest *IQ* obtained. Fillmore also correlated item performance in baby tests with later *IQ*.

Bayley's correlations for 61 children between the performance on her series of tests at 4-5-6 and 7-8-9 months and performance at 18-21-24 and 27-30-36 months ranged from +.10 to +.39. Muhlenbein correlated a quotient obtained from the Linfert-Hierholzer tests at six months with the *IQ* at four and five years, obtaining a coefficient of +.11. This correlation was raised to +.33± 08 when sigma scores for the early test and a restricted age range for the later tests were used. Fillmore's correlation between performance on the scale at 6 months and the earliest *IQ* (at two to three years) was +.320. In addition to the correlation of the total test score with the earliest *IQ* she correlated certain of her items with this criterion, obtaining coefficients for the six months items ranging from +.02 to +.43.

These investigations, though not dealing with the six months Gesell schedule as a unit, do deal with many of the items of the Gesell schedule and probably sample the Gesell series rather adequately. It is obvious that there is little correlation shown in these investigations between performance at six months and later test performance This lack of correlation may be due to a number of factors, including poor reliability of the tests, influence of environment on the abilities measured, or a too gross lumping of specific factors which may be common to various age levels but the measurement of which is submerged by improper weighting.

The object of this report is to analyze the predictive efficiency not only of the total test at six months but of items within the test

As stated above, a number of our children were given the Merrill-Palmer test and the 1916 Stanford Revision of the Binet test at later age levels. For correlation with future test performance the items and the total test scores at six months on the Gesell series were correlated with the Merrill-Palmer at 24 months (on 48 children), and the Binet at 36 months (on 31 children). Total Gesell scores at six months were also correlated with the Merrill-Palmer at 18 months, but with a group of only 20 subjects

The intercorrelations between tests were as follows:

	N	*r*
Item score Gesell 6 months and Merrill-Palmer raw score 18 months	20	+.62±.09
Item score Gesell 6 months and Merrill-Palmer raw score 24 months	48	+.37±.08
Item score Gesell 6 months and Stanford-Binet mental age 36 months	31	+.46±.10
Raw score Merrill-Palmer 18 months and raw score 24 months	13	+.62±.12
Raw score Merrill-Palmer 18 months and Stanford-Binet mental age 36 (too few cases)		
Raw score Merrill-Palmer 24 months and Stanford Binet mental age 36	29	+.66±.07

It will be seen that relationships are nowhere unusually high. Correlations between adjacent age periods,—6, 18, 24, and 36 months—are about .65, and highest in the group. The Merrill-Palmer at 24 months correlates as well with the Binet at 36 as it does with the 18 months Merrill-Palmer. It is interesting to note that the Gesell series at six months correlates slightly better with the Binet at 36 months than it does with the Merrill-Palmer test at 24 months It will be seen later that the mean item-correlations with the 24 months Merrill-Palmer and the 36 months Binet bear out this tendency quite markedly

For predictive purposes these coefficients are, of course, extremely low, presenting coefficients of alienation for the Gesell series and other tests of .78 for the 18 months Merrill-Palmer, .88 for the three year Stanford-Binet, and .94 for the two year Merrill-Palmer. Indices of forecasting efficiency are as low as 6 per cent for the two year Merrill-Palmer. However, it will be seen that the coefficients are higher than those reported by other investigators

CORRELATIONS BETWEEN PERFORMANCE ON TEST ITEMS OF THE SIX MONTHS GESELL SERIES AND PERFORMANCE ON THE MERRILL-PALMER TEST AT TWENTY-FOUR MONTHS

In the hope of determining whether or not an analysis of relationships by items might indicate a higher predictive value for certain of them than for the total six-months series en masse, it was decided to correlate each of the items passed by at least 90 per cent of the group at six months with the performance at 24 months (Merrill-Palmer raw score) and at 36 months (Stanford Binet Mental Age). It should be kept in mind that the three year Binet on 31 subjects correlated slightly better with the Gesell six months series than did the two year Merrill-Palmer on 48 subjects

TABLE 9

Score	Frequency
6 to 8	3
9 to 11	0
12 to 14	4
15 to 17	11
18 to 20	11
21 to 23	8
24 to 26	6
27 to 29	3
30 to 32	2
Total	48
Mean	19.88
Sigma	5.57

The distribution of raw scores for the Merrill-Palmer test at two years is shown in Table 9

Item correlations were determined by means of the biserial coefficient technique The coefficients secured, in order of their size, appear in Table 10

TABLE 10

Regards pellet	58
Picks up cube	49
Splashes in tub	44
*Looks for fallen object	43
Music laughs	40
Reaches directly for spoon	39
Throws objects to floor	39
*Lifts inverted cup	38
Holds two cubes	37
Prefers one hand	37
Sits with slight support	36
Casts objects for noise	33
Dangling ring: clasps ring	29
Dangling ring seizes above head	28
Dangling ring persistent reaching	27
Drops one cube for third	.27
Exploratory manipulations	25
Music stops crying	25
Takes bottle in and out of mouth	22
Conscious of fallen object	22
*Pellet whole hand	19
Makes stepping movements	15
Paper purposeful reaction	14
Bangs spoon	11
Secures cube (from inverted cup)	08
Creeps or hitches	.07
Reacts to mirror image	04
Inhibits head and one hand	03
Rolls back to stomach	03
Pats table	00
Says "mama" and "dada"	.00
*Sits alone	—05
Manipulates one hand	—07

*Fillmore coefficients with earliest *IQ*

Looks for fallen object	30
Hunts covered object	22
Picks up pellet	43
Sits alone	25

The average of these coefficients, at 23, is much lower than the correlation between the total Gesell Schedule and the Merrill-Palmer test, which was 33. However, certain items correlate much better than did the total test, namely, the first 11 items Interestingly enough, many of these are not items of outstanding motor character. The items *regards-pellet, splashes in tub, looks for fallen objects,* and *music: laughs* are not as "motor" as some of the other items.

As one might expect, item correlations tend to be lower in the case of the 24 months Merrill-Palmer than in the case of the total test at six months.

It will be seen that the ranking of six-months items as they correlate with the score on the Merrill-Palmer at two years is quite different from the ranking according to their correlation with the total six months test Again the postural-locomotor items rank low, and the one language item ranks very low Items which tended to correlate low with the Gesell at six months but highly with the Merrill-Palmer at 24 months were

Regards pellet
Splashes in tub
Music: laughs
Holds two cubes
Prefers one hand
Clasps dangling ring
Music stops crying

Those which correlated highly with the Gesell at 6 months, but low with the Merrill-Palmer were

Secures cube from inverted cup
Reacts to mirror image
Inhibits head and one hand
Pats table
Manipulates one hand

There is a tendency for items which we have designated as "Awareness or distance perceptive" to increase in their relative ranking, while the "motor" items tend to decrease. The mean coefficients for the groups compared in the preceding analysis of relationships to the total test, in this case were, for the motor items, 25, for the awareness-distance perceptive .32, the posture-locomotor .09, and for the one language item 00. Whereas the motor items, perhaps due to their greater abundance in the 6 months schedule tended to correlate best with the total number of items passed, when correlated with the 24 month Merrill-Palmer they do not correlate as well as do the awareness or distance perceptive items. Posturo-locomotor items correlate poorly with six months as with two year total performance.

Considering the items in terms of their predictive value for Merrill-Palmer performance, it is obvious that though this value is considerably higher in some cases than for the total test, it is yet low. For the highest coefficient, of 58, the coefficient of alienation is .81, and the forecasting efficiency but 19 per cent.

The method of multiple correlation was tried in several ways to determine whether or not higher correlation and hence higher predictive values might be obtained. The items which correlated with the Merrill-Palmer test at 24 months better than did the total 6 months score (or better than .33), and which at least 25 and not over 75 per cent passed, were used in these correlations. These items, and their zero order coefficients are given in Table 11.

TABLE 11

	Biserial r	Coefficient alienation	Predictive efficiency
Regards pellet	58	81	19%
Splashes in tub	.44	90	10%
Looks for fallen object	43	90	10%
Music, laughs	.40	.92	8%
Reaches directly for spoon	39	92	8%
Throws objects to floor	39	.92	8%

Correlating the Merrill-Palmer at two years with combinations of two items gave only five coefficients higher than that for *regards pellet* above. These all involved the item *regards pellet,* and are given in Table 12.

TABLE 12

	R	Coefficient alienation	Predictive efficiency*
Regards pellet combined with	66	75	25%
Looks for fallen object	66	75	25%
Music laughs	.64	77	23%
Splashes in tub	60	.80	20%
Reaches directly for spoon	60	.80	20%
Throw objects to floor	.59	.81	19%

*Formula for "Predictive Efficiency," $100[1-(1-r^2)]$ or $100(1-k)$

These coefficients represent very little increase in predictive efficiency as the result of combining items in pairs

Multiples were obtained in another way by adding as independent variables each of the above items in the decreasing order of their zero order coefficients. This effect is shown in Table 13

TABLE 13

Item added to those above it	Its zero *order r*	Cumulative *R*	Predictive *efficiency*
Regards pellet	58		19%
Plus splashes in tub	.44	60	20%
Plus looks for fallen object	43	66	25%
Plus music laughs	.40	.67	26%
Plus reaches directly for spoon	39	67	26%
Plus throws objects to floor	39	.70	29%

It is apparent that the addition of more and more items increases the correlation and the predictive efficiency; indeed, it would seem that were we to continue this process we might get a fair degree of correlation. On the other hand, these items are those which correlate best with the criterion, and they exhibit no greater correlation between them than do other items in the scale. We might regard the coefficient of 70 as somewhere near the maximum correlation obtainable with the raw score of the Merrill-Palmer at two years. This degree of correlation is but 10 per cent more efficient for prediction than is the single item, *regards pellet*.

A general conclusion from the relationships here demonstrated is that the Gesell Schedule at six months is not greatly predictive of mental status as measured by the Merrill-Palmer test at two years. Certain items in the scale at six months are relatively more predictive, but their correlation is not over .60. In no sense may we regard the six months Gesell as measuring more than 50 per cent of whatever abilities the Merrill-Palmer measures at two years.

CORRELATIONS BETWEEN PERFORMANCE ON TEST ITEMS OF THE SIX MONTHS GESELL SERIES AND PERFORMANCE ON THE STANFORD-BINET (1916 Revision) AT THIRTY-SIX MONTHS

Correlations as made for the 24 months Merrill-Palmer were also made for the Stanford-Binet at 36 months. The 1916 Revision was used, and correlations were made with mental age. Distribution of mental ages at three years is given in Table 14.

It will be seen that the distribution is definitely bimodal, with peaks at 37 months (*IQ* about 100) and again at 47 months (*IQ* of 135). This distribution and the small number of cases obtained at this level might decrease the significance of statistical treatment as compared with the relationships estimated for the Merrill-Palmer

TABLE 14

MA	Frequency
34 to 36	6
37 to 39	8
40 to 42	2
43 to 45	3
46 to 48	7
49 to 51	3
52 to 54	1
55 to 57	1
Total	31
Mean	42.95
Sigma	5 97

test at two years Despite these handicaps it would seem to us interesting, for purposes of comparison between the Binet at three and the Merrill-Palmer test at two years, to relate to them early test performance on the Gesell items

It has been mentioned that the raw score on the six months Gesell correlated +.47 with Binet mental age at three years

The biserial technique of correlation was used for the Binet relationships to Gesell items at six months. Coefficients secured for each item, in decreasing order, are given in Table 15.

Correlations of six-months items with the Stanford-Binet at three years ranged from .00 to .58, grouping about 20- 25. The highest coefficient, with *splashes in tub* was obtained with an item correlating very low with the total test. The following items correlated low with the total test but well with the Binet at 36 months

Music. laughs
Splashes in tub
Regards pellet
Prefers one hand
Music stops crying
Casts objects for noise
Sits alone
Pellet whole hand
Says "Mama" or "dada"

Those correlating low with the Binet, but highly with the total test were.

Looks for fallen object
Paper purposeful reaction
Throws objects to floor
Lifts inverted cup
Picks up cube

TABLE 15

Splashes in tub	.58
Regards pellet	.56
Reaches directly for spoon	.54
Prefers one hand	53
Music: stops crying	.52
Music: laughs	.52
Dangling ring: persistent reaching	.46
*Sits alone	.46
Casts objects for noise	44
Inhibits head and one hand	36
Exploratory manipulations	.36
Dangling ring: seizes above head	.36
Pats table	.34
*Pellet: whole hand	32
Secures cube (from inverted cup)	.31
Says "mama" and "dada"	.30
Reacts to mirror image	.30
Takes bottle in and out of mouth	.29
Bangs spoon	.26
Drops one cube for third	.25
Conscious of fallen object	23
Manipulates one hand	.20
Picks up cube	.18
*Lifts inverted cup	.18
Makes stepping movements	.17
Rolls back to stomach	.16
Throws objects to floor	.10
*Looks for fallen object	07
Paper: purposeful reaction	.02
Creeps or hitches	.02
Holds two cubes	.00

*Fillmore coefficients with earliest *IQ*:

Sits unsupported	.25
Picks up pellet	43
Hunts covered object	.22
Looks for fallen object	.30

As in the case of the 24 months Merrill-Palmer correlations, the strictly motor items tend to have lower coefficients than the awareness or distance perceptive items. The postural-locomotor items remain low throughout. The lone language item correlates slightly, for the first time, with the Binet. Interestingly enough, *sitting alone* which correlates poorly with all other tests, correlates relatively higher with the Binet.

Table 16 presents the average correlation for the motor, awareness, posturo-locomotor and language items with each of the total test scores.

TABLE 16

Type	Number items	Gesell 6 mos Total test less item	Merrill-Palmer 6 mos.	Stanford-Binet 36 mos
Motor	21	.65	25	.30
Awareness	7	59	32	.41
Posturo-Locomotor	5	.55	11	21
Language	1	.45	.00	.30
Mean		60	.23	30
Correlation total test		.88	.33	.47

It will be seen above that the Gesell series at six months consists to a large degree of motor items These involve reaching, manipulation, grasping, etc , and naturally correlate well with the total test due to sheer abundance. They do not correlate as well with either the Merrill-Palmer at two years or the Binet at three years, however, as do items which we have designated as "awareness" items These awareness items were *reaches for dangling ring,*[8] *reaches directly for spoon,*[8] *reacts to mirror image, looks for fallen object, regards pellet, music. laughs, conscious of fallen objects, throws objects for noise,*[8] and *music· stops crying* They were set aside in a group because (with the exception of three items designated as fitting into the motor series) they did not seem primarily to involve manipulation of any sort, nor response to tactile stimulus. Essentially they involve reaction to a somewhat removed source of stimulus, although the response itself is of motor character. In a sense they might be defined as perceptive responses, or responses which would indicate reaction to a specific stimulus. By contrast with other items they imply a stimulus differential; they are not qualities of performance.

In an effort to see whether or not the predictive value of items in the Gesell series for three year Binet mental age might be raised by multiple correlation, a technique similar to that used for the Merrill-Palmer relationships was used Five items were selected which were passed by from 25 to 75 per cent of the group, and which correlated more highly with the Binet at three than did the total six months test (which correlated 47). These, with their coefficients of correlation and alienation, and their indices of predictive efficiency are given in Table 17.

---

[8] Also included under motor items.

TABLE 17

Item	Biserial r	Coefficient alienation	Index forecasting efficiency
Splashes in tub	58	81	19%
Regards pellet	.56	83	17%
Reaches directly for spoon	54	84	16%
Music: stops crying	.52	.85	15%
Music: laughs	52	85	15%

Correlating combinations of two items with the Binet mental age at three gave the coefficients shown in Table 18.

TABLE 18

Combination	*R*	Coeff alien	Index F.E
Regards pellet and music, stops crying	.76	65	35%
Splash and music laughs	75	66	34%
Splash and reach for spoon	.71	70	30%
Splash and music stops crying	.70	72	28%
Regards pellet and music, laughs	69	72	28%
Splashes and regards pellet	.65	76	24%
Regards pellet and reach spoon	.65	.76	24%
Reach spoon and music stop crying	64	.77	23%
Reach spoon and music laughs	.62	79	21%

These coefficients represent an increase in predictive efficiency of almost double that of the zero order correlations

When multiple coefficients were obtained by adding as independent variables each of the above items in order of their decreasing zero order coefficients, the results were obtained as shown in Table 19

TABLE 19

Item added to those above it	Its zero order *r*	Cumulative *R*	Predictive efficiency
Splashes in tub	.58		19%
Plus *regards pellet*	.56	65	24%
Plus reaches directly for spoon	54	73	31%
Plus music stops crying	.52	79	39%
Plus music laughs	52	.78	37%

The results (Table 19) would suggest that the predictive efficiency of single items is better than that for the total test, and is raised by combining items. When five items correlating relatively highly with the Binet at three years, at about 54 are combined, a correlation of almost .80 is obtainable, which has a predictive efficiency of almost 40 per cent If we regard the abilities tested by the Gesell series at six months as basic to those measured by the Binet at three years, we may conclude that as much as 60 per cent of the variance in mental age at the later age level is sampled at six months.

SUMMARY

A group of 123 infants were given the Gesell Schedule of items for the six months level These children ranged in age from five to eight months No correlation was found between number of items passed and age Despite this fact, further correlations were limited to the group of 77 children tested within four days of the sixth month birthday. In this group no significant sex differences were apparent, either for the raw score or between the percentages passing single items.

The difficulty of the items ranged from 100 per cent passing to 4 per cent passing Weighting items for their difficulty gave scores which correlated almost perfectly with the number-of-items-passed score, so was considered as no more advantageous

Item analysis showed that items involving reaching, grasping, and manipulation correlated most highly with the total score at six months; this was due to their abundance in the list of 46 items Items involving distance perception and awareness, posturo-locomotor abilities, and the one language item correlated but slightly with the total test The central tendency for item correlation with the total test was about .66 (lowered to about 60 when considered as correlation with the total test less the item correlated).

The literature is in agreement in reporting six months performance to correlate poorly (from .10 to 30) with indices of later mental status. Our own analysis showed the relationship of total test score to two year Merrill-Palmer raw score to be .37, and to the three year Stanford-Binet mental age 46 Single items combined in certain ways gave multiple coefficients of as great as .70 for the two year test and 80 for the three year test

When the nature of the six months items was considered in terms of their correlation with test performance at two and three years, there was found a tendency for manipulative items to decrease and for items of distance perception and awareness to increase in relative importance.

Compared with other studies, this report indicates that the Fels group may be somewhat superior in ability to other groups (with the possible exception of the Iowa Fillmore group). This analysis also indicates a greater degree of correlation between performance at six months of age and later test performance. To explain this increased correlation is difficult. Low correlation between mental test performance at various age levels might be accounted for, as Muhlenbein (8) has indicated, by a number of factors. Low reliability of the tests used, lack of constancy in the actual abilities measured or modification of them by environmental forces, the degree to which ability is adequately sampled by a test at a given age, or an actual replacement of abilities in the repertoire at one level by a different set of abilities at another: these and other possible explanations might be adduced to account for low correlation. If these are the chief factors, then, in a situation of augmented correlation, we must assume that they are less operative. The reliability of tests used would not seem to be different in this study from what it is in others. The fact that we have used the Gesell Schedule as Gesell described it *might* make for a wider sampling of abilities at six months than would the items selected and added to by Bayley, Fillmore, and Linfert and Hierholzer, though this is difficult to demonstrate in any way. It is possible also that had these investigators used multiple correlation they might have obtained coefficients of a size better approximating our coefficients.

Having considered these factors, there remains the possibility that at six months, and at two and three years, there is a common matrix of abilities which may be sampled at each point Certain abilities are common to each of the tests, though they form but a small portion of what the tests cover,—never over 50 or 60 per cent, and probably less These abilities are modified, of course, by environment. It seems possible that the range of environmental influences may be less in the case of the Fels group than in the case of the other groups, and that with this limitation in range there is a tendency for environmental factors which make for enhancement of abilities to stand out, at the sacrifice of those most detrimental. In other words, the Fels environments are of limited range, and are relatively more conducive to mental growth than those of other studies This would be in agreement with the findings of Muhlenbein (8).

An important next step in the analysis of abilities at early stages is to determine the nature of factors which make for performance. At present we are preparing a report on the results of factor analysis of Gesell items at the six months level, and it is hoped that we may do a similar analysis of this schedule at 12, 18 and 24 months. However, such an approach should require caution, since we are not sure that factor analysis will tell us validly more than do the simpler techniques here used. Nevertheless, it may be possible when this is done, to speak more clearly of the essential abilities involved in performance at these early levels, and eventually to relate them to specific factors in later development on the one hand, and to newborn and even fetal behavior on the other

REFERENCES

1 BAYLEY, NANCY Mental growth during the first three years. A developmental study of sixty-one children by repeated tests *Genet. Psychol Monog*, 1933, **14**, 1-92

2 ——— The California first year mental scale. Berkeley, Calif Univ. California Press, 1933

3 ———. The development of motor abilities during the first three years *Monog Soc Res. Child. Develop*, 1935, No. 1, 1-26.

4 FILLMORE, EVA A Iowa tests for young children *Univ Iowa Stud. Child Welfare*, 1936, **11**, 1-58 Vol XI, No 4, 1-58

5 GESELL, A. The Mental Growth of the Pre-school Child A psychological outline of normal development from fifth to the sixth year, including a system of developmental diagnosis New York· Macmillan, 1925 Pp x+447.

6 JAFFA, A S The California pre-school mental scale. Univ. Calif Press, Syllabus Series No 251, 1934 Pp 66.

7. LINFERT, HARRIETTE-ELISE, & HIERHOLZER, HELEN M. A scale for measuring the mental development of infants during the first year of life *Stud Psychol & Psychiat*, 1925, **1**, No. 4. Pp v+33

8 MUHLENBEIN, JOSEPHINE N The validation of the Hierholzer Scale. Unpublished Master's Thesis, Catholic Univ America, Washington, D C

9 STUTSMAN, RACHEL Mental measurement of pre-school children with a guide for the administration of the Merrill-Palmer Scale of mental tests Yonkers-on-Hudson, N Y World Book Co, 1931 Pp x+368.

10 TERMAN, L M The Measurement of Intelligence An explanation of and complete guide for the use of the Stanford Revision and extension of the Binet-Simon intelligence scale Boston Houghton Mifflin, 1916 Pp. xx+362

Samuel S Fels Research Institute
Antioch College
Yellow Springs, Ohio

The Journal of Genetic Psychology, 1938, 52, 327-331.

STUDIES IN MENTAL DEVELOPMENT II. ANALYSIS OF ABILITIES TESTED AT THE AGE OF SIX MONTHS BY THE GESELL SCHEDULE*

Samuel S Fels Research Institute, Antioch College

T. W RICHARDS AND VIRGINIA LAFAYETTE NELSON

In the preceding report in this series (1) it was shown that the Gesell schedule of six months items seemed to consist of an abundance of "motor" items involving reaching, grasping, and manipulating which as single items correlated well with the total test at that level, but not highly with mental ability at two and at three years Indeed, certain items which correlated poorly with the total test at six months were superior to these motor items in their correlation with later mental ability It was tentatively indicated that the nature of these latter items was somewhat that of alertness on the one hand, and distance perception on the other.

Because of the indicisive quality of these findings with respect to the Gesell items at six months, it was decided to intercorrelate certain of the tests The items selected for intercorrelation were those items passed by the middle 50 per cent of children from 25 to 75 per cent Tetrachoric coefficients were calculated between each of 17 of these items with each other For these correlations 77 subjects were used, in all cases the distribution represented at least 60 subjects

The items and their correlations are presented in Table 1 It will be seen that all coefficients are positive, and range from 02 to 1 00 There is a fair degree of correlation throughout, the coefficients average about .48 Sigmas for these coefficients varied from 00 to 21

Thurstone's method of multiple factor analysis was used with the coefficients presented in Table 1

Space will not permit presentation of the tables of residuals after extraction of factors Table 2 will indicate the decline in the mean and standard deviation of the residuals

It will be seen that after the second matrix (from which the third

---

*Received in the Editorial Office on September 3, 1937

TABLE 1

INTERCORRELATIONS (TETACHORIC) BETWEEN SEVENTEEN ITEMS AT SIX MONTHS

1 Splash	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1 Splashes in tub		30	.15	.25	21	.17	28	.29	.30	30	.14	.04	53	06	- 12	30	08
2 Pats table			85	59	77	69	.39	.31	.42	03	69	.66	38	.67	60	80	.15
3 Inhibits head and hand				.60	96	.83	.40	.52	66	.07	80	59	49	82	64	1 00	36
4 Reaches direction spoon					46	.48	51	.58	58	38	.73	44	44	66	.58	69	47
5 Manipulates hand						68	34	.51	51	05	70	.43	.52	68	55	1 00	.34
6 Makes stepping movements							.40	30	62	17	.57	41	31	65	.42	61	20
7 Looks for fallen object								.73	.44	.45	.60	35	21	.50	47	.52	75
8 Throws object to floor									44	32	62	47	51	46	.35	62	61
9 Takes bottle in and out mouth										56	67	21	29	.44	30	40	58
10 Music stops crying											.23	05	02	08	.24	25	89
11 Exploratory manipulations												50	52	68	58	.75	09
12 Bangs spoon													25	53	.41	59	22
13 Regards pellet														38	.46	58	.23
14 Dangling ring, persist. reaching															69	.73	.48
15 Drops one cube for third																51	.50
16 Secures cube																	36

TABLE 2

Matrix	Mean residual	Standard deviation	Extremes
Correlation	478	215	02 to 1,00
First	—006	.148	—41 to 63
Second	—006	093	—35 to 28
Third	—008	083	—17 to 22
Fourth	—010	074	—22 to 16
Fifth	—006	063	—20 to 18

factor loadings are extracted) not only the middle two thirds, but practically all the residuals are within the area defined by ± the sigma (about 16) of the average original correlation coefficient Further, it will be seen that after the third factor was extracted there appeared to be no definite change in either central tendency or variation of residuals; they seem after this point to vary at random within a range roughly of 15 to —20

It was decided on the basis of the results indicated in the foregoing paragraph to stop the analysis after three factors had been extracted.

The factor matrix obtained from the analysis is shown in Table 3 A number of rotations[1] were tried, the clearest delineation of

TABLE 3

Item	I	II	III	Uniqueness
Splashes in tub	34	—20	35	.72
Pats table	78	43	13	.20
Inhibits head and one hand	.91	.44	—05	00
Reaches directly for spoon	78	—09	01	39
Manipulates one hand	82	40	—16	14
Makes stepping movements	.71	.23	.06	44
Looks for fallen object	.69	—39	—.24	32
Throws objects to floor	71	—27	10	41
Takes bottle in and out of mouth	69	—27	.18	43
Music, stops crying	42	—75	00	27
Exploratory manipulations	82	20	—24	23
Bangs spoon	58	.23	.22	56
Regards pellet	57	08	—37	53
Reaches for dangling ring	.79	26	.31	.21
Drops one cubs for third	69	12	.26	45
Secures cubs from inverted cup	91	28	—.22	05
Music laughs	61	—66	—29	11

---

[1] For the benefit of those who might wish to experiment with the factor matrix, we might indicate that the rotations made were of I and II 29°, I' and III 40°; I'' and II' 23°, I''' and III' 16°, II'' and III'' 18°; II''' and I'''' 12°, II'''' and III''' 15°; to give I'''', II'''' and III''''

axes appearing to us to yield the arrangement (excluding values less than 20) as shown in Table 4.

TABLE 4

Item	I	III	II
Reaches for dangling ring	54	69	
Pats table	.71	54	
Bangs spoon	42	50	
Inhibits head and one hand	89	.45	
Manipulates one hand	86	31	
Drops one cube for third	41	58	23
Makes stepping movements	.58	43	20
*Splashes in tub		45	28
Reaches directly for spoon	46	38	.51
*Takes bottle in and out of mouth	22	46	56
Throws objects to floor	27	40	59
Secures cube	89	29	28
Exploratory manipulations	78	22	.35
Regards pellet	60		.33
Looks for fallen object	34		75
*Music: stops crying			83
*Music laughs			.93

*Items reported by mother

Factor 1 seems to be common to all tests *excepting those which require the mother's report* If we disregard Factor 1 for the moment and examine Factors II and III it would appear that the former is most present in items which are less motor, more of the "awareness" type described in the preceding paper. Factor III is most prevalent in more obviously motor-manipulative-reaching-grasping items, and not at all present in the "awareness" items It seemed logical to name Factor II *alertness* and Factor III *motor ability* (to cover manipulating, grasping, reaching, stepping). But what of Factor I?

In naming Factor I there seem to us to be three possibilities for consideration. In the first place, if we exclude the tests which depend on the report of the mother, it is seen that the factor is common to all other tests in some abundance. It is present generally in the greatest degree in those tests which, as we have shown earlier (1) tend to correlate highly with the total test performance at six months. If the factor is a general factor applicable to all except the mother's-report items, *it cannot in any sense be regarded as "g" or intelli-*

gence, since the tests in which it is in least abundance correlate best with intelligence at two and three years The opposite is also true: Factor I is most present in tests which correlate poorly with intelligence at two and three Again, were the Factor a "*g*" factor we might expect it to be present in the items reported by the mother.

A second possibility is to consider Factor I as a factor of "mental set" which appears (in varying degree) in the testing situation but not in the behavior noted by the mother and reported to the examiner This would suggest the possibility that the infant on the day of a particular examination is more (or less) testable,—that is, he is functioning generally at a higher or lower level on a given day This tendency characterizes in most abundance those items which most represent the whole test—the manipulative-reaching-grasping items.

The third possibility is that Factor I represents a "halo effect" characteristic of the examination but not of the mother's-report items, which again is most effective on items of the manipulative-reaching-grasping character, as well as most characteristic of what the test as a whole seems to tap.

From this brief report the following conclusions may be made

1. Three factors seem to account for most of the variance in behavior of the six months old infant as tested by 17 items in the middle range of difficulty. These items represent a fair sampling of behavior at this level, but may have an abundance of "motor" as opposed to non-motor items.

2 The three factors may be designated as (*a*) "*testability*" or "*halo effect,*" (*b*) "*alertness*" and (*c*) "*motor ability*"

Reference

1 Nelson, Virginia Lafayette, & Richards, T W Studies in mental development 1 Performance on Gesell items at six months and its predictive value for performance on mental tests at two and three years *J Genet Psychol*, 1938, 52, 303-325

Samuel S Fels Research Institute
Antioch College
Yellow Springs, Ohio

The Journal of Genetic Psychology, 1938, 52, 333-346

THE USE OF PRONOUNS BY YOUNG CHILDREN: A NOTE ON THE DEVELOPMENT OF SELF-AWARENESS*

Institute of Child Welfare, University of Minnesota

FLORENCE L GOODENOUGH

Because pronouns are more highly generalized symbols than other parts of speech, their early use by children has considerable psychological significance as a means of throwing light on the development of conceptual thinking To some extent, this fact has been recognized One of the standard questions that concerned the earlier students of child psychology had to do with the process of learning to distinguish between the self and the rest of the universe The use of the first personal pronoun has commonly been taken as evidence that at least a primitive stage in the development of self-awareness has been reached, and child biographers have therefore been at pains to note the age at which this pronoun first becomes a part of the child's spoken vocabulary. Surprisingly enough, however, few of the published records contain an adequate account of the conditions under which the word was first used, and since most of the reports are concerned only with a single child, few conclusions of general significance are warranted.

The study about to be reported is based upon observational records of the spontaneous conversations of children enrolled in the nursery school and experimental kindergarten of the University of Minnesota, Institute of Child Welfare. The method employed by McCarthy (2) which consists in recording for each child a sample of 50 consecutive responses under uniform conditions was employed.

In McCarthy's study, the child under observation was alone in a room with an adult observer who provided him with toys and picture books but avoided direct verbal stimulation as far as was consistent with maintaining an easy natural atmosphere. Our study consists of two sets of records for each child The first was secured under conditions identical with those reported by McCarthy, except that we used a small testing room in the nursery school building

---

*Received in the Editorial Office on September 14, 1937

whereas McCarthy's subjects were observed in their own homes The second set of records was taken during the free-play hour in the nursery school or kindergarten. Series I (hereinafter known as the "Controlled" or *C Series*) thus includes remarks made to an adult or undirected monologue during solitary play with toys, while Series II, called the "Free" or *F Series*, consists chiefly of remarks made to other children, although, inasmuch as every verbal response was recorded from the time of beginning the observation until 50 such remarks had been secured, the records also include some remarks made to teachers and a fair amount of monologue not clearly directed to any particular person

Each child was observed individually within a month of the date midway between birthdays As a rule, the two observations for a given child were made on consecutive days; the interval between observations was never greater than a week. All observations were made by persons accustomed to the speech of children; they are for that reason likely to be somewhat more accurate than would be the case had they been taken by individuals of less experience in this field While independent observers were found to agree very closely in their records, there is reason to think that this agreement understates the true accuracy of recording, since it is not always easy for two full-sized adults to occupy the most favorable position for observation at the same time. However, a preliminary experiment indicates that under condition *C*, the correlation between the total number of pronouns reported for each of 16 children by two simultaneous observers is approximately 96 when age is held constant while under condition *F*, where the maintenance of equally favorable conditions of observation is more difficult, the corresponding figure is around .92 It was not feasible to secure a sufficient amount of data to determine conclusively whether the agreement between observers varies with the age of the children observed, but examination of the discrepancies in the records as well as common-sense judgment suggests that such a relationship exists and this for two reasons, first, the greater difficulty of understanding the speech of the younger children and, secondly, the relatively small number of pronouns used at those ages, and the consequent low standard deviations of the individual scores.

This study differs from most others on the development of language in children, since it is concerned mainly with the changes in the content of children's speech under different types of social stimulation. Most of the previous studies have been based upon only a single situation, or in those instances where the speech in two or more situations has been compared, the number of cases has been small or the composition of the groups has varied from situation to situation. Smith (6) reports a study of two sets of records taken under conditions similar to those of the present study, but there is much uncontrolled variation in the conditions of study. The records were obtained from several American nursery schools and kindergartens by several different investigators, and from white children in Hawaii While it is stated that some of the subjects were observed in both situations, there is no information as to the extent to which this was done or as to the usual lapse of time between observations in these cases Since the number of observations under the controlled situation was 107 as compared to 198 during free-play; it is evident that the groups are far from being identical The major part of the report, moreover, is concerned with age and sex differences, the situational analysis is very cursory. In an unpublished study by McConnon, (4) the language responses of nursery school children were studied under a number of different conditions, but the group used was not large and because of fluctuating conditions of attendance, the composition of the various subgroups varies slightly from one situation to another. The number of cases also varies slightly but is usually fewer than 25 In none of these studies was any special analysis of pronoun usage attempted.

McCarthy (3) observed each of 31 nursery school children in two situations similar to those reported here. This study was carefully controlled but because of the small number of cases at each age, conclusions were of necessity very tentative However, two trends of probable significance were noted, viz., more than three times as many emotionally-toned responses and a greater tendency toward egocentric responses during free-play than in the controlled situation. Of the 31 children observed, 29 gave more egocentric responses in Situation *F* than in Situation *C*.

In classifying a given response as egocentric or socialized, McCarthy used a modification of Piaget's functional method. This procedure is admittedly rather subjective, although it has the advantage of being based on complete units of speech rather than upon single words In the present study, we have approached the ques-

TABLE 1

MEAN FREQUENCY OF OCCURENCE OF CERTAIN SPECIFIED GROUPS OF PRONOUNS PER SAMPLE OF 50 RESPONSES, BY AGE, SEX, AND CONDITIONS OF OBSERVATION

Age and sex	Boys 2½		Girls 2½		Boys 3½		Girls 3½		Boys 4½		Girls 4½		Boys 5½		Girls 5½	
Condition	C	F	C	F	C	F	C	F	C	F	C	F	C	F	C	F
No. cases	9		9		17		21		18		25		55		49	
Pronouns																
I. I, me, myself	3.1	3.0	2.8	9.2	10.4	10.9	7.9	14.2	9.2	11.4	11.4	16.1	11.2	12.9	11.3	15.4
II. My, mine	3.3	4.3	3.7	1.7	0.8	2.3	1.0	4.0	1.1	1.9	0.8	2.5	1.1	2.1	1.3	2.8
III We, us, ours	0.8	—	0.6	—	1.2	1.4	1.5	1.8	1.4	3.3	1.2	2.4	0.8	1.6	1.8	1.9
IV. You, your, yours, yourself	0.4	0.5	0.6	1.9	4.1	4.7	2.5	6.9	3.7	6.9	4.0	8.0	2.9	5.9	3.7	7.5
V. He, him, his, she, her, hers, himself, herself	0.8	0.5	0.9	0.7	4.2	0.7	3.7	2.1	4.3	2.2	3.9	2.6	4.2	2.4	5.4	1.5
VI They, them, their, theirs (persons)	—	—	—	—	0.4	0.1	0.8	—	0.6	0.6	0.4	1.2	0.3	0.3	0.7	0.3
VII Who, whose	—	—	—	—	0.7	—	0.3	0.3	0.6	0.5	0.4	0.2	0.2	0.1	0.4	0.4
VIII It, that, its, what, this	12.4	1.6	10.8	3.9	11.7	5.7	11.5	7.9	9.4	7.2	12.5	9.4	11.0	10.2	11.8	10.2
IX. They, them, those, these, their, theirs (objects)	13.8	—	—	1.6	4.8	2.1	4.8	2.4	5.8	2.3	6.0	3.4	6.0	3.3	5.6	3.3
Total pronouns	24.6	9.9	19.4	19.0	38.3	28.9	34.0	39.6	36.1	36.3	40.6	45.8	37.7	38.8	42.0	43.3

tion of egocentricity by means of an analysis of the pronouns used in the two situations This method has the merit of objectivity since the only source of error lies in the recording of the original responses and this, as was stated before, is probably a small one. Moreover, the type of analysis made here has been useful in bringing to light certain trends in the development of conceptual thinking that have hitherto escaped observation

Table 1 shows the mean number of occurrences of certain specified pronouns in each sample of 50 consecutive remarks by a total of 203 children under the two situations described.

Examination of Table 1 reveals interesting variations in the use of certain specified pronouns or pronoun-groups according to the age and sex of the children and the conditions under which the observations were made. For both sexes and at all ages with the single exception of the small group of two-year-old boys, pronouns of the first person singular occur with far greater frequency during free play with other children than in the controlled situation where the child presumably feels less need to assert himself. The same trend is shown in the use of the possessions, *my* and *mine*. Insofar as the use of these pronouns is indicative of something in the nature of an ego-consciousness, it is evident that this feeling is brought to the fore in the more competitive situations of group play far more frequently than is the case during the less socialized conditions of the controlled situation. This is wholly in accordance with McCarthy's findings

Strictly speaking, the first personal pronoun can have no true plural The so-called plural forms (*we, us, ours*) involve an idea of cooperation or of identifying oneself with the group that seems to be relatively late in appearing. Among the two-year-olds there is no record of the use of any of these words in the free-play situations, and the few scattering instances of their use with adults suggest a copying of adult phraseology rather than a true concept of joint action. It may be noted in this connection that in Shirley's study of 25 babies (5) only a single instance of the use of the first person plural is recorded. After the age of two, these pronouns are used somewhat more often in the *F* than in the *C* situation, but the frequency is not great at any age.

Pronouns of the second person, like those of the first, occur much more frequently during play with other children than when alone with an adult. They are not often used by the two-year-olds. We have not attempted to distinguish between the singular and the plural usage since the identity of word-form makes differentiation on the basis of the written record impossible in the majority of cases, and even at the time of observation, the distinction is not always evident

In classifying pronouns of the third person we have attempted, as far as possible, to distinguish between those referring to persons and those referring to objects and here we have again separated the singular from the plural forms Examination of Table 1 shows that each of these groups of pronouns shows a marked excess in frequency of use in the *C* situation, and that the plural forms referring to persons are used but rarely at these ages. Shirley records a single instance of the use of the word "*them*" by a child under the age of two, but does not specify whether the reference was to persons or to objects.

The use of the neuter pronouns, *it, that,* etc , greatly exceeds the use of the more specific masculine and feminine forms at all ages, and this is especially true of the plurals. Moreover, the use of the neuter pronouns is much more frequent in the controlled situation where the child is playing with toys and looking at pictures than it is in the free situation where other children are present and competing for his attention. The explanation for this is so obvious that the point would hardly seem worth mentioning except for the fact that proportions of the various parts of speech in the language of children are frequently quoted in a fashion that might lead the reader to think that these proportions depend almost wholly upon the child's level of development and vary but little with the conditions of observation. The unsoundness of this point of view will be apparent from an examination of the data here presented. As a matter of fact, while most of the pronouns here listed show a distinct difference in frequency of usage according to the immediate situation under which they are used, age differences are not always apparent, even when no allowances are made for the increasing length of the sentence. Considering absolute frequency, the pronouns that show greatest increase in usage with advancing age are those of the first and second persons and to a somewhat less extent, the third person singular when referring to persons and not to objects. The neuter pronouns show little absolute change in frequency as age advances

This appears to be due to the fact that the younger child characteristically uses these pronouns as a substitute for the names of unknown objects "What's *that*? *That's* a funny thing" "*It* makes a noise, doesn't it?" etc, are expressions heard over and over again in any nursery school

The comparatively rare use of third-person plurals referring to persons is another point of interest Apparently, for the young child at least, objects lend themselves to grouping much more readily than do people Our data do not show a single instance of the use of third-person plural pronouns with personal antecedents before the age of 3½, after which the typical frequency is roughly one occurrence in two samples, that is, about once in 100 responses.

Whether or not third-personal pronouns referring to persons are equally rare in the speech of older children and adults is not easy to determine because of the difficulty of reproducing the situations used in this study My guess would be that adults use them more frequently than children do, though probably not so often as the same pronouns with non-personal antecedents As a partial and admittedly very inadequate check, I counted the number of such pronouns in ten random samples of written discourse, each from a different book and a different writer and each sample a full page in length The books included fiction [4], biography [1], psychology [1], anthropology [1], sociology [2], and popular science [1]. The number of sentences in each sample varied from 9 to 21, the total number was 131 The number of third-person plural pronouns referring to persons ranged from 0 to 6; the total number was 17, while those not referring to persons number 24 The total number of words was 3420, or an average of 26.1 words per response, which is approximately 5½ times the average length of the sentences used by the five-year-old-girls in the controlled situation Thus it appears that in these samples of written discourse by adults, pronouns of the third-person plural that refer to persons occurred about 2½ times as frequently as in the spoken discourse of preschool children while those not referring to persons are found only about 1/3 as often [1] Whether we are dealing here with a true age difference or whether the apparent difference is purely the result of the different character of the material used cannot be said with assurance from the data at hand, but scrutiny of the original records of the children's speech shows many instances in which the names of individual children are enumerated in detail when an adult would almost certainly have used one of the collective pronouns, *they* or *them*.

---

[1] These ratios are computed for samples containing an equal number of words. They are approximations only, since reference to Table 1 will show that there is a good deal of shifting about from one group to the next in the total frequency of these pronouns We have taken 0 5 and 5 0 as

Thus far we have considered changes with age in the use of pronouns only in an absolute sense, without regard to the fact that sentence-length is also increasing with age. Table 2 shows the percentages obtained by dividing the mean number of pronouns of each type by the mean number of words in the sample

When allowance is made for the increase with age in the typical length of the responses (see last line of Table 2) the proportion of pronouns of all classes shows no consistent trend with age in the *C* situation, while in the *F* situation there is for the boys a slight indication of an increase in the proportion of pronouns as age advances. This tendency does not hold for the girls, however, and as the differences shown by the boys are small, they may represent only a chance departure from a stable ratio Likewise, there is no consistent difference between the two situations in the total percentage of pronouns used.

Certain of the pronoun groups, however, show interesting variations with age and condition of experiment. The increase with age in frequency of use of the first personal pronoun, holds good for the relative as well as for the absolute comparison. The greater use of these pronouns during free play shows up even more clearly in terms of the percentage of pronouns than in their absolute frequency, since the average length of sentence is shorter in the *F* than in the *C* situation. The same may be said of pronouns of the second person. Third person singular neuters make up a steadily decreasing proportion of the total number of words in the *C* situation, but the percentages tend to increase with age in the *F* situation.

In all probability it is this preponderance of the third person

---

the most representative figures for the frequency of pronouns of the first and second type respectively in a sample of 50 responses totalling 238 words, which is the average for five-year-old girls in the *C* situation. This is compared with 17 of the first and 24 of the second in an adult sample of 3420 words

TABLE 2

RATIO OF THE NUMBER OF PRONOUNS OF VARIOUS CLASSES TO THE TOTAL NUMBER IN THE SAMPLE

(In percentages)

	Boys 2½		Girls 2½		Boys 3½		Girls 3½		Boys 4½		Girls 4½		Boys 5½		Girls 5½	
					C	F	C	F	C	F	C	F	C	F	C	F
I, me, etc.*	2.6	2.6	2.5	8.7	5.9	7.2	4.4	10.8	4.5	6.1	5.2	8.2	5.8	6.8	5.1	8.0
My, mine	2.8	3.8	3.3	1.6	0.5	1.5	0.6	3.1	0.5	1.0	0.4	1.3	0.6	1.1	0.6	1.5
We, etc.	0.7	—	0.5	—	0.7	0.9	0.8	1.4	0.7	1.8	0.5	1.2	0.4	0.8	0.8	1.0
You, etc	0.3	0.4	0.5	1.8	2.3	3.1	1.4	5.3	1.8	3.7	1.8	4.1	1.5	3.1	1.7	3.9
He, she, etc.	0.7	0.4	0.8	0.6	2.4	0.5	2.1	1.6	2.1	1.2	1.8	1.3	2.2	1.3	2.4	0.8
They, (pers.) etc.	—	—	—	—	0.2	0.1	0.4	—	0.3	0.3	0.2	0.6	0.2	0.2	0.3	0.2
Who, etc	—	—	—	—	0.4	—	0.2	0.2	0.3	0.3	0.2	0.1	0.1	0.1	0.2	0.2
It, etc.	10.4	1.4	9.8	3.7	6.6	3.8	6.4	6.0	4.6	3.8	5.7	4.8	5.7	5.4	5.3	5.3
They (non-pers.) etc.	3.2	—	—	1.5	2.7	1.4	2.7	1.8	2.9	1.2	2.7	1.7	3.1	1.8	2.5	1.7
Total	20.7	8.6	17.4	17.9	21.7	18.5	19.0	30.2	17.7	19.4	18.5	23.3	19.6	20.6	18.9	22.6
Mean no. words in sample	118.8	114.0	110.6	106.1	177.1	151.0	179.3	131.1	203.3	187.2	222.0	195.8	191.7	188.5	222.8	191.5
Mean no. words per response	2.37	2.28	2.20	2.13	3.34	3.02	3.58	2.62	4.06	3.75	4.44	3.91	3.83	3.77	4.45	3.83

*For a complete list of the pronouns included in each group, see Table 1 neuter singulars among the pronouns used in the *C* situation that accounts for the fact that, contrary to popular opinion, the absolute number of pronouns used in the course of 50 responses shows no increase with age after the maximum has been reached at 3½ years, and in proportion to the total number of words used the number of pronouns actually declines. The use of an unusually large number of pronouns in this situation may thus be regarded as evidence of linguistic immaturity, rather than maturity in children over the age of 3½ years.

Table 3 shows the correlations between the total number of pronouns in the odd-numbered responses and the corresponding number in the even-numbered responses by age and sexes separately after correction for total length by means of the Spearman-Brown prophecy formula Although these correlations are not extremely high, they indicate a degree of stability of the data that is sufficient for group comparisons. In view of the small number of pronouns belonging to the various sub-classes in the individual records, it has not seemed worth while to compute self-correlations for this part of the analysis. It is apparent from the data given in Table 1 that a sample of 50 responses is not sufficient for the study of individual differences in the use of separate pronouns, but the consistency of the group trends shown in Tables 1 and 2 is evidence of their validity as group measures

TABLE 3

ODD-EVEN CORRELATION OF TOTAL NUMBER OF PRONOUNS IN SAMPLE CORRECTED BY SPEARMAN-BROWN PROPHECY FORMULA

	2½		3½		4½		5½	
	B	G	B	G	B	G	B	G
Free	87	95	72	70	.68	88	77	.58
Controlled	85	40	77	91	.89	.81	.76	.86
No cases	9	9	17	21	18	25	55	49

That females are somewhat more precocious than males in the use of language is demonstrated by their earlier mean age of beginning to speak, the use of longer sentences during the preschool years, larger vocabularies during childhood, and higher scores on reading tests and English composition tests at all ages McCarthy has shown that in all available measures of speech in which age progression appears, girls are somewhat more advanced than boys during the preschool years. Studies of college students have shown that in many respects, at least, this superiority of the female sex is maintained throughout the adolescent period. Our data show a similar trend with respect to the use of pronouns. The mean total number of pronouns in the *F* sample in greater for girls than for boys at every age. In the *C* situation, however, the total pronoun count for boys is greater than that for girls at the two younger ages. Examination of Table 1 provides a possible explanation. At these ages the third person singular neuters account for more than half of the pronouns used in that situation. The superior vocabularies of the girls enable them to substitute actual names of objects in many instances where the boys still cling to the more primitive "this" and "that." Thus the greater total number of pronouns used by the boys at these ages should probably be interpreted as an indication of linguistic inferiority rather than superiority.

The ratio of the number of first personal pronouns to the number of second and third combined has by some writers been taken to have some significance as an indication of egocentricity. A study by Anderson (1) in which the data consisted of samples of compositions written by students in a junior college, reports a slightly higher average value for this index in the compositions written by women students than in those written by men, but the difference obtained is smaller than its standard error and so may be only a chance variation. For the children of our groups, no consistent sex difference in respect to this tendency was established, but the proportion of first personal pronouns used in the *F* situation greatly exceeds that found in the *C* situation. This is true, even when the ratio is based only upon the pronouns referring to persons with all those having a neuter antecedent (both singulars and plurals) excluded (see Table 4).

TABLE 4

RATIO OF THE NUMBER OF FIRST PERSON SINGULAR PRONOUNS TO THE NUMBER OF THOSE IN THE SECOND AND THIRD PERSON

	A				B			
	C		F		C		F	
Age	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
2½	0.37	0.53	2.81	1.34	5.67	4.46	7.30	4.26
3½	0.43	0.38	0.99	0.92	1.28	1.28	2.41	2.03
4½	0.42	0.45	0.67	0.75	1.19	1.48	1.37	1.58
5½	0.42	0.45	0.78	0.78	1.65	1.29	1.75	1.97

A All pronouns including neuters.
B Only pronouns referring to persons; neuters omitted

Of the 16 sex comparisons in Table 4, the pronoun-index is higher for the boys than for the girls in 7 instances, the girls lead in 7 cases and in the remaining cases the ratios are equal. It is interesting to note that age progression is much more clearly seen in the free play situation than it is in the more rigidly controlled but less highly socialized condition of the experimental room, and that under the former conditions both sexes show a fairly rapid drop in this ratio from the age of 2½ to 4½ with a slight rise at the age of 5½. Although this suggested increase in self-assertiveness among the five-year-olds may be only a chance variation from a stabilized value it is nevertheless not wholly without interest, in view of recent findings by Wolf (7) and others which suggest that social competition first becomes clearly apparent in the average child around the age of five years. The fact that the five-year-olds were enrolled in kindergarten while the other children were all in the nursery school should also not be overlooked, since, as has been pointed out before, the immediate situation is so important a factor in determining the form and content of children's speech.

Two facts of decided importance for students of the development of speech in childhood have been brought out by this study First, it has been shown that the immediate situation exercises an important effect upon the form and content of speech, and that accordingly, any statements regarding the average or usual characteristics of the language of persons of any age must be limited to the conditions under which the records were secured or to such other situations as have been empirically demonstrated to be their psychological equivalents Secondly, it is clear that many of the formal grammatical classifications of adult usage are inadequate to bring out significant developmental trends in the speech of children. The very marked changes that occur in the use of pronouns, for example, are almost wholly obscured when all pronouns are grouped into a single class. Developmental processes are qualitative as well as quantitative, and in devising systems for classifying behavioral manifestations it is necessary to keep these qualitative changes in mind if the systems are to be useful It is suggested that in the study of children's language, too much attention has been paid to the type of grammatical analysis used by adults and too little to the developmental changes in conceptual thinking and social drives that lie back of the verbal expression In this respect, Piaget has set an example by which many of us can well profit.

SUMMARY

1 Samples of the spontaneous conversations of 203 children in two psychologically dissimilar situations have been analyzed with respect to the frequency of usage of certain specified pronouns in each situation Each sample consisted of 50 consecutive responses

2. In agreement with the findings of other investigators, the percentage of pronouns in the total sample shows little consistent change with age or sex after the age of three years Certain specified groups of pronouns, however, show very pronounced changes with age and with the conditions of observation.

3. Pronouns of the first person singular (including the possessives) are used far more frequently during play with other children than when the child is alone with an adult Both singular and plural pronouns of the third person with non-personal antecedents show the opposite trend in very marked degree.

4. Pronouns of the third person plural with personal antecedents are but rarely used at these ages

5. Neuter pronouns of the third person singular make up a decreasing proportion of the total number of words used as age advances This is due to the fact that the young child characteristically uses these pronouns in place of the names of unfamiliar objects. As vocabulary increases, nouns are gradually substituted for the indefinite neuters.

6. Sex differences are not large but tend to favor the girls in those aspects of pronoun usage that show a developmental trend. An index of egocentricity obtained by dividing the number of first personal pronouns by the sum of those in the second and third persons was unrelated to sex, but showed a high relation to the immediate situation

REFERENCES

1 ANDERSON, J E An evaluation of various indices of linguistic development *Child Develop.,* 1937, **8**, 62-68.

2 MCCARTHY, DOROTHEA The Language Development of the Preschool Child. University of Minnesota Institute of Child Welfare Monograph Series No 4 Minneapolis: Univ Minnesota Press, 1930 Pp xiv+174.

3 ——— A comparison of children's language in different situations and its relation to personality traits. *Ped. Sem & J Genet Psychol*, 1929, **36**, 583-591

4. McCONNON, KATHLEEN C. The situational factors in the language responses of nursery school children. Unpublished Ph.D. thesis, Institute of Child Welfare, University of Minnesota, 1935.

5. SHIRLEY, MARY. The First Two Years: A Study of Twenty-Five Babies. Intellectual Development. University of Minnesota Institute of Child Welfare Monograph Series No. 7. Minneapolis: Univ. Minnesota Press, 1933. Pp. xvi+513.

6. SMITH, MADORAH E. A study of some factors influencing the development of the sentence in preschool children. *J. Genet. Psychol.*, 1935, **46**, 182-212.

7. WOLF, THETA HOLMES. The effect of praise and competition on the persisting behavior of kindergarten children. To be published shortly as No. 16 in the University of Minnesota, Institute of Child Welfare Monograph Series.

Institute of Child Welfare
University of Minnesota
Minneapolis, Minnesota

The Journal of Genetic Psychology, 1938, 52, 347-360

THE RELATION OF MENTAL STAMINA TO PARENTAL PROTECTION*

Department of Psychology, Northwestern University

JOHN J B MORGAN AND MARY H. BANKER

Most observers of human behavior will agree that there are wide differences in the manner in which individuals react to a problem situation At one extreme is the person who is stimulated by a difficulty, who meets the challenge of a problem with increased activity, emotional excitement, and increased mental alertness At the other extreme is the person who wilts at the slightest opposition, who slinks away from a problem, whines for help, complains about his hardships, and seeks some means of escape Between these two extremes can be found all graduations This variable is probably not a unitary trait and is probably dependent for its strength upon a number of different factors.

This paper reports an attempt to gather some systematic evidence as to the operation of parental protection in producing a change in a child's response to problem situations. Does a child who has been overprotected by his parents show less mental stamina than does a child who has been privileged to use his own resources and to solve his own problems? Will a child tend to "give up" more readily in the face of obstacles if he has been overprotected by his parents, or will such overprotection train him to fight the harder to overcome difficulties?

Theoretical discussions and clinical observations have led to the hypothesis that a child must attain a balance between dependence upon elders and his development of independence. He has a need for security—a feeling of being wanted, of belonging, or being accepted and completely welcome On the other hand, with increasing maturity he has a need for freedom—the need to be an individual in his world, to contribute to it, and to be respected

The parental overprotection which may influence this balance between dependence and independence includes such activities on the part of the parent as babying, oversolicitude, too much fondling,

---

*Received in the Editorial Office on September 21, 1937

prolongation of infantile care, prevention of independent behavior, excessive supervision, too much negative restraint, and the like.

In order to study this problem of the relation between the two variables, it is necessary to have some measurement of the degree to which a child is stimulated by a difficult situation as well as a means of measuring the degree of protection given the child by his parents.

To measure the degree to which a child will respond to a challenge, the Morgan-Hull (1) persistence maze was used. The principles upon which this maze is based may be briefly summarized:

1 The first of any new series of problems should be so easy that the individual can readily solve the problem and thus gain some self-assurance.

2. This first success should be followed by problems of increasing difficulty, but with continued success with each problem.

3. The difficulty of succeeding problems should be graded so as to provide an increasing challenge to the individual. Eventually he should be given a very difficult problem.

4. The individual's final response should not be measured in terms of success, nor in terms of the length of time he persists in his attempted solutions, but in the attitude which he manifests toward the task

The Morgan-Hull maze fulfills these requirements. It is a stylus maze with the starting box in the center. The entire maze is covered with a screen which moves with the stylus and which permits exposure of a very small segment of the maze adjoining the stylus. Four problems of increasing difficulty are used in succession with each subject (Figure 1). In Problem 1, a barrier is placed at (1) and the home pocket is located at *A* This cuts off most of the maze and offers a very simple problem to the subject. He is permitted to repeat this maze until he makes two successful runs In Problem 2, a barrier is set at each of the points marked (2); one of these cuts off part of the maze and the other one closes the home pocket at *A*. The home pocket for this problem is at *B*. Thus, the subject has to unlearn his first problem and to learn one somewhat more difficult. In Problem 3, barriers are set at the three points marked (3), thus leaving open the home pocket *C* and blocking of some of the maze. In Problem 4, the home pockets *A*, *B*, and *C* are all closed and the entire maze is opened. This last problem can-

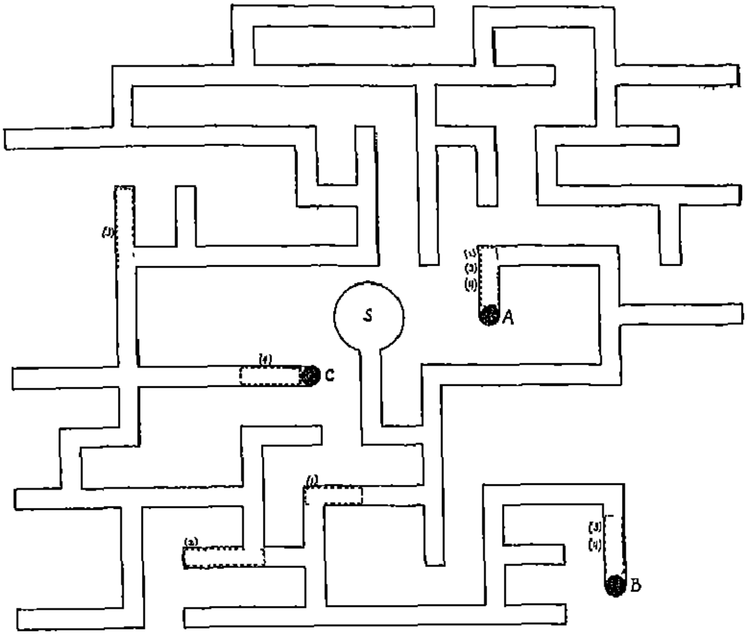


FIGURE 1
THE MORGAN-HULL PERSISTENCE MAZE

not be solved—there are no home pockets open. Nevertheless, the subject is led to believe that the maze can be solved because of his previous successes with Problems 1, 2, and 3.

The subject's performance in this final unsolvable problem is rated by two independent observers on the following nine-point scale.

1 Careless,—anxious to quit the task.

2. Excuse hunter—readily gives some excuse to get out of the task—feels badly, eyes hurt, time is valuable, etc.

3. Fiddling plodder—keeps working because he apparently has not enough initiative to try harder or to quit—follows the line of least resistance

4. Intermittent worker—goes by spurts, working hard and then having periods of fiddling.

5 Works hard, but has little insight. Works hard but with

little intelligence. Never suspects that the maze cannot be worked. Works in a blind fashion.

6. Persistent worker with some method or definite attempt to reach the goal.

7. Persistent worker with some insight. Probably tries two or three different methods of reaching goal but shows increasing discouragement.

8. Tenacious, obstinate type—more determined to succeed because of the obstacles. Failure acts as a challenge—the greater the difficulty the harder he works.

9. Analytical type—intelligently persistent to the extent that he fully analyzes the problem. Presents data or reasons why he thinks that the maze cannot be solved.

Although this scale is qualitative, different subjects can be rated by means of it with considerable accuracy.

To measure the degree of protection afforded to the child in his home, a multiple choice and true-false test was devised by Miss Banker. The first five items were not used in the scoring, but were given to enable the examiner to eliminate those children who were not living with father or mother.

The instructions given to the children taking the test were:

> There are no right or wrong answers to these questions except as they are right or wrong for each of you individually. What is right for you may be wrong for your neighbor. You are to answer every question as truthfully as possible. Your answers will be absolutely secret. Neither your classmates, your teacher, your parents, nor any other person you know will ever see your paper. Please raise your hand if there is anything you do not understand. Be sure to answer every question.

The scoring was determined by averaging the judgment of several psychologists as to the significance of each possible answer. The numbers in parenthesis preceding each answer give the weight assigned to that answer.

THE PROTECTION TEST

Draw a line under the true answer

1 Yes No My mother is living.
2 Yes No My father is living.
3 Yes No I live with my father but not with my mother.
4 Yes No I live with my mother but not with my father.

5. Yes No I live with my father and my mother
6 (1) Yes (3) No Outside of school hours I can do as I please
7 (3) Yes (0) No When I entertain friends in my home, my mother helps me
8 (1) Yes (3) No I am given spending money regularly
9 (2) Yes (4) No I can spend my allowance any way I want without telling anyone what I spend it for
10 (0) Yes (3) No I am allowed to entertain friends in my home.
11 (2) Yes (3) No I am allowed to choose my own clothes.
12 (0) Yes (3) No I can accept invitations to my friends' parties without getting permission from home
13 (2) Yes (1) No As a small child, I was whipped when I misbehaved
14 (0) Yes (4) No If a pilot asked me to go for an airplane ride, my parents would let me go
15 (1) Yes (0) No I share a room with my mother and father
16 (1) Yes (0) No My parents go to bed when I do, usually
17 (0) Yes (0) No My grandmother often lets me do things that have been refused me by my parents
18 (0) Yes (1) No My parents pay me for work done for them
19 (1) Yes (0) No When I am sick my parents pay me to take medicine
20 (2) Yes (0) No My parents pay me to "be good."
21 (2) Yes (0) No I am sick more than anyone else in the family

Put a check mark in the space in front of the true answer

22 My father treats me
  (3) (a) as if I were younger than I am
  (0) (b) as if I were older than I am
  (0) (c) as if I were just my own age
23 My mother treats me
  (3) (a) as if I were younger than I am
  (0) (b) as if I were older than I am
  (0) (c) as if I were just my own age
24 Commands given me by my father
  (2) (a) are always willingly obeyed
  (0) (b) are usually willingly obeyed
  (0) (c) are always willingly obeyed when they seem reasonable
  (1) (d) are seldom obeyed when they seem unreasonable to me
  (3) (e) always have to be obeyed even though they seem unreasonable to me.
  (3) (f) are never obeyed unless I feel like it
25 Commands given by my mother
  (2) (a) are always willingly obeyed
  (0) (b) are usually willingly obeyed
  (0) (c) are always willingly obeyed when they seem reasonable
  (1) (d) are seldom obeyed when they seem unreasonable to me

- (3) (e) . always have to be obeyed even though them seem unreasonable to me
- (3) (f) .. . are never obeyed unless I feel like it.

26 When I earn any money
- (2) (a) . .I give it to my mother or father
- (2) (b) . .I put it in the bank or save it
- (4) (c) .. my parents tell me what to buy with it
- (1) (d). .. I spend it as I please

27 When I want to do something very badly and my parents won't let me I can get my own way if I
- (3) (a). cry.
- (3) (b). pout
- (3) (c) ..beg for it long enough
- (3) (d) .. get sick
- (3) (e) .. break my toys
- (3) (f) run away
- (3) (g) .....refuse to eat
- (3) (h). I can't get my own way no matter what I do

28 When my mother scolds me, my father
- (2) (a).. always sides with my mother.
- (0) (b). never takes sides
- (1) (c). ..occasionally takes my part
- (2) (d) .frequently takes my part.

29 When my father scolds me, my mother
- (2) (a). always sides with my father.
- (0) (b) . never takes sides.
- (1) (c) . . .occasionally takes my part.
- (2) (d) . frequently takes my part.

30. If I break a rule that my mother wants me to observe, it is
- (4) (a) . never overlooked
- (2) (b) occasionally overlooked.
- (2) (c) ... seldom permitted to go unnoticed
- (1) (d) frequently unnoticed.

31 If I break any rule that my father wants me to observe, it is
- (4) (a) never overlooked.
- (2) (b) occasionally overlooked
- (2) (c) . .seldom permitted to go unnoticed
- (1) (d) .. . frequently unnoticed

32 When my mother tells me to do or not to do a certain thing, she expects me to obey
- (1) (a) . ..after being told why
- (2) (b) . . .without questioning why
- (3) (c) . .instantly

33 When my father tells me to do or not to do a certain thing, he expects me to obey
(1) (a) . after being told why
(2) (b) . without questioning why
(3) (c) instantly

34. I feel that my mother disciplines me
(0) (a) fairly.
(1) (b) . mildly
(2) (c) in a lax manner
(3) (d) unfairly.
(2) (e) severely
(3) (f) in a strict manner

35 I feel that my father disciplines me
(0) (a) fairly.
(1) (b) mildly.
(2) (c) in a lax manner
(3) (d) . unfairly
(2) (e) severely.
(3) (f) in a strict manner

36 I have to be at home
(2) (a) at a time agreed upon when I leave home
(2) (b) . . at a certain hour on school nights only
(6) (c) at a certain hour every night.
(0) (d) at no special time

37 In the summer, I am allowed to go swimming
(4) (a) . if my mother or my father goes with me
(0) (b) with boys and girls my own age
(0) (c) . . by myself
(6) (d) I am not allowed to go swimming at all.

38 When I tell my parents why I do certain things, my reasons are
(0) (a) always accepted
(0) (b) usually accepted.
(1) (c) . usually not accepted
(1) (d) sometimes doubted
(3) (e) . . always doubted

39. My parents try to improve me
(0) (a) by pointing out my faults
(0) (b) by showing me my good and bad points
(2) (c) . by threatening punishment
(2) (d) . by punishing me

40 When I do something my parents have told me not to do, my privileges are
(3) (a) always taken away
(2) (b) . usually to some extent taken away

(1) (c) . . seldom taken away
(0) (d). . . . never taken away.

41. The friends I go out with, the places I visit, the friends I invite to my home
(5) (a) . . must be approved by my mother
(4) (b) . . must be approved by my father
(3) (c) . . must be approved always by someone
(0) (d) . need not be approved by anyone

42 I usually go to the movies or theater
(4) (a) with my mother
(4) (b) . with my father
(1) (c) . with my brother or sister.
(0) (d) with my best friend
(0) (e). . with a group of friends
(0) (f) by myself
(5) (g) I am not allowed to go at all

43 My mother
(0) (a) pays little attention to what I do
(2) (b) usually attends to the important things I do
(6) (c) directs everything I do

44. My father
(0) (a) . pays little attention to what I do.
(2) (b) . usually attends to the important things I do.
(6) (c) directs everything I do

45. Whenever I have problems or troubles, my father
(2) (a) always takes time to listen to me
(3) (b) . always helps me straighten them out.
(1) (c) seldom helps me with them
(1) (d) . is always too busy to help me with them
(0) (e) . . . never listens to me

46. Whenever I have problems or troubles, my mother
(2) (a) . always takes time to listen to me
(3) (b). always helps me straighten them out.
(1) (c) seldom helps me with them
(1) (d) is always too busy to help me with them
(0) (e). . never listens to me

47. When I find myself doing a job which is harder than I thought it would be
(0) (a) I try harder to get it
(2) (b) . . I get my parents to help me with it
(0) (c) . I decide that it is too much for me and do something else

48. I usually tell my secrets to
(2) (a) . . my mother.
(2) (b) my father

(0) (c) my brother or sister
(0) (d) my best friend
(0) (e) no one

49 When I'm out riding my bicycle
(2) (a) my mother worries for fear I'll get hurt
(4) (b) I am allowed to go only a certain distance from home
(0) (c) I can ride anywhere I please
(3) (d) my parents won't let me have a bicycle because they are afraid I'll get hurt
(0) (e) I have no bicycle but *not* for the reason that my parents are afraid I'll get hurt

50 At home, when meals are served
(1) (a) it doesn't matter whether I am late or not
(2) (b) I am always required to be on time
(0) (c) our meals are not served at exact times

51 My parents let me decide important things for myself
(0) (a) usually
(0) (b) frequently
(3) (c) seldom
(5) (d) never

52 My mother "preaches" or "lectures" to me about how I behave
(5) (a) constantly
(3) (b) frequently
(0) (c) occasionally
(0) (d) seldom
(0) (e) never

53 My father "preaches" or "lectures" to me about how I behave
(5) (a) constantly
(3) (b) frequently
(0) (c) occasionally
(0) (d) seldom
(0) (e) never

54 I am whipped
(3) (a) frequently
(2) (b) occasionally
(1) (c) seldom
(0) (d) never

55 I am punished in some way when I am naughty
(3) (a) always
(2) (b) usually
(1) (c) seldom
(0) (d) never

56 I am allowed to explain my conduct before being punished
(0) (a) always

(1) (b) usually
(2) (c) seldom
(3) (d) never.

57 I deserve the punishment I get
(1) (a) always
(1) (b) usually.
(2) (c) seldom.
(3) (d) never.

58. If my father refuses something that I have asked for, my mother grants it
(3) (a) frequently.
(2) (b) usually.
(1) (c) occasionally.
(0) (d) never.

59 If my mother refuses something that I have asked for, my father grants it
(3) (a) frequently.
(2) (b) usually.
(1) (c) occasionally
(0) (d) never.

60 I may bring my friends home with me without getting permission first
(0) (a) always
(1) (b) usually.
(2) (c) seldom
(3) (d) never.

61 I have to take care of my younger brothers and sisters
(3) (a) constantly.
(2) (b) frequently
(1) (c) occasionally
(0) (d) seldom
(0) (e) never
(0) (f) I have no younger brothers or sisters

62 I am allowed to go to the movies alone
(0) (a) always.
(0) (b) frequently.
(0) (c) occasionally
(3) (d) seldom
(5) (e) never.

63 My mother helps me with my home work when I have trouble with it
(3) (a) always.
(2) (b) usually.
(1) (c) seldom
(0) (d) never.

64 My father helps me with my home work when I have trouble with it
(3) (a) always
(2) (b) usually
(1) (c) seldom
(0) (d) never

65 My mother worries about me
(3) (a) constantly
(1) (b) at times
(0) (c) never

66 My father worries about me
(3) (a) constantly.
(1) (b) at times
(0) (c) never

67 My mother gives me whatever I ask for
(3) (a) always
(2) (b) usually
(1) (c) seldom
(2) (d) never

68 My father gives me whatever I ask for
(3) (a) always.
(2) (b) usually
(1) (c) seldom
(2) (d) never.

69 My mother asks me to stay home and keep her company
(4) (a) constantly
(2) (b) usually
(0) (c) seldom
(0) (d) never.

70 My father tells me I have to stay home
(3) (a) constantly
(2) (b) occasionally
(1) (c) seldom
(0) (d) never

71 My mother gives up things for herself so that I can have the things I want This happens
(3) (a) constantly
(1) (b) occasionally
(0) (c) seldom.
(0) (d) never

72 My father gives up things for himself so that I can have the things I want. This happens
(3) (a) constantly
(1) (b) occasionally

(0) (c) seldom
(0) (d) never

73 My older brothers or sisters have to take care of me
(3) (a) constantly
(2) (b) frequently
(1) (c) occasionally
(0) (d) seldom.
(0) (e) never
(0) (f) I have no older brothers or sisters

To obtain suitable subjects for the experiment, the children in the seventh and eighth grades in one school were given the test for parental protection. After excluding all those who were not living with both parents or who did not fill in the examination properly, 51 good papers remained. These 51 children were then given the Morgan-Hull persistence maze test and their performance rated independently by two observers. In addition these same children were given the Rogers *Personality Schedule,* the Morgan *Mental Test,* and were rated by the teachers of both the seventh and eighth grades as to their mental stamina. These ratings were based on the same scale that was used by the experimenters in scoring the children upon their performance of the Morgan-Hull persistence maze, but the teachers based their judgments upon the everyday contact with the children and not upon the children's specific response to any set task. The results of these various tests and ratings are given in Table 1.

These data reveal Pearson's Product-Moment correlation coefficients as given in Table 2

Between parental protection as measured by the test devised by Miss Banker and mental stamina as measured by the maze persistence test there is a correlation of —.80. This is equivalent to a coefficient of alienation of .60, that is, the ability to predict the persistence of a child in the maze test from a knowledge of his record on the protection test is 40 per cent better than chance. In other words, the degree of protection given in the home (as measured by the protection test) accounts for 64 per cent of the factors determining the performance in the persistence maze test. The ability to predict mental stamina from intelligence is only 4.6 per cent better than chance as indicated by the correlation coefficient of .32 and a similar influence is shown with the correlation of .33 with the Rogers' Personality score. The striking factor is the total lack of correlation between teachers' ratings and parental protection and the persistence

TABLE 1

Case	Protection	Protection repeated	Maze	7th Grade teachers' R	8th Grade teachers' R	Average teachers' R	Mental test	Rogers total	Rogers inferiority	Rogers social mal	Rogers fam mal	Rogers daydreaming	Age	Siblings
1	119	99	3	5	5	5	36	51	22	10	17	2	13	3
2	118	93	1	5	6	5.5	83	61	14	24	16	7	13	5
3	110	91	1	6	6	6	68	24	13	6	4	1	13	1
4	110	99	3	2	6	4	45	37	13	14	6	4	12	0
5	104	93	4	9	9	9	80	35	16	7	8	4	11	4
6	103	85	4	8	6	7	54	36	18	13	3	2	12	1
7	101	103	5	6	6	6	64	45	17	18	7	3	13	2
8	100	97	3	5	5	5	71	36	15	7	14	0	13	1
9	97	93	3	9	9	9	101	34	7	14	11	2	11	2
10	96	68	2	6	4	5	77	34	11	13	10	0	13	1
11	96	81	4	2	5	3.5	41	48	14	24	7	3	12	2
12	91	83	3	1	3	2	51	28	9	8	5	6	13	5
13	90	94	4	6	6	6	89	24	8	7	7	2	13	2
14	87	98	5	6	6	6	81	36	10	16	10	0	13	1
15	86	75	1	9	6	7.5	45	32	13	14	5	0	11	2
16	86	75	5	1	5	3	46	35	14	18	12	1	13	1
17	85	82	3	4	4	4	62	24	10	1	10	3	13	1
18	85	71	5	6	6	6	63	36	13	13	8	2	13	2
19	85	73	5	2	5	3.5	45	40	17	13	6	4	11	1
20	84	85	5	6	6	6	68	46	16	17	6	7	13	2
21	83	81	1	8	6	7	72	40	19	7	10	4	12	5
22	83	89	5	8	6	7	75	29	11	14	4	0	11	1
23	83	75	5	7	6	6.5	83	32	8	16	8	0	11	2
24	82	57	5	4	4	4	93	38	11	16	6	5	13	0
25	81	85	5	8	6	7	85	31	7	16	8	0	13	3
26	81	81	5	2	4	3	81	56	22	17	13	4	13	2
27	81	70	5	5	6	5.5	48	24	7	8	9	0	12	3
28	81	60	5	6	6	6	85	37	16	12	9	0	11	4
29	80	74	5	2	4	3	68	50	11	11	18	10	13	4
30	80	83	5	2	7	4.5	69	29	4	18	5	2	12	1
31	80	70	6	6	4	5	35	38	17	15	6	0	12	1
32	79	79	6	6	6	6	67	36	10	16	8	2	12	1
33	79	83	6	7	9	8	115	49	4	23	17	5	12	2
34	79	79	6	2	1	1.5	67	27	9	13	5	0	13	0
35	78	72	7	8	9	8.5	101	28	4	11	3	0	14	1
36	78	80	7	6	6	6	59	27	11	10	6	0	11	0
37	77	90	6	3	4	3.5	56	40	16	13	7	4	12	1
38	77	81	5	8	6	7	85	23	5	11	6	1	11	2
39	76	81	7	3	6	4.5	77	35	7	20	8	0	13	1
40	75	63	7	6	9	7.5	97	24	9	11	4	0	12	0
41	74	52	8	7	6	6.5	61	39	19	14	6	0	12	1
42	72	81	8	9	4	6.5	83	39	13	16	6	4	11	1
43	72	76	9	4	6	5	54	35	12	16	7	0	13	1
44	69	66	7	6	4	5	97	26	15	5	6	0	13	3
45	68	93	6	3	1	2	54	28	8	10	8	2	12	1
46	67	70	8	6	6	6	51	30	8	14	4	4	11	0
47	65	69	8	7	6	6.5	83	30	9	9	8	4	12	1
48	62	45	9	6	6	6	94	30	6	13	9	2	13	3
49	60	74	8	6	6	6	79	49	18	17	8	6	13	2
50	60	65	7	8	6	7	74	30	12	10	6	2	11	1
51	56	73	7	4	6	5	84	27	11	7	5	4	13	3

TABLE 2

Parental Protection with Persistence Maze	— 80± 03
Parental Protection with Ave Teachers' ratings	02± 09
Parental Protection with Intelligence	— 02± 09
Parental protection with Roger's—total score	34± 08
Parental Protection with Personal Inferiority	34±.08
Parental Protection with Social Maladjustment	.08±.09
Parental Protection with Family Maladjustment	30± 09
Parental Protection with Daydreaming	— 31± 09
Persistence Maze with Average Teachers' ratings	.02±.09
Persistence Maze with Intelligence	.32± 08
Persistence Maze with Rogers' total score	33± 08
Teachers' Ratings with Intelligence	48± 07
Teachers' Ratings with Rogers' total score	—.13± 09
Intelligence with Rogers' total score	— 08± 09
Parental Protection I with Parental Protection II	64±.05
Teachers' Rating I with Teachers' Rating II	53± 07

maze score Teachers' ratings correlate highest with the intelligence test scores It seems apparent that the teacher tends to judge the persistence or mental stamina of a child in terms of his intelligent behavior, and not in terms of his ability to be stimulated or challenged by a problem situation.

CONCLUSION

There is a significant relationship between a child's tendency to react to a problem situation with increased zeal as the difficulty of the problem is increased, and a lack of excessive supervision and domination by the child's parents The kind of stamina that a child manifests in attacking a difficult situation is not the sort of behavior that teachers are inclined to judge as persistence or stamina. They are more influenced by the child's intelligent behavior than by his resistence to thwarting.

REFERENCE

1 MORGAN, J. J. B, & HULL, LUCILE The measurement of persistence *J. Appl Psychold*, 1926, 10, 180-187.

Department of Psychology
Northwestern University
Evanston, Illinois

The Journal of Genetic Psychology, 1938, **52**, 361-374

THE FREE WORD ASSOCIATION OF ELEMENTARY SCHOOL CHILDREN: II VERBAL RESPONSES*

Department of Psychology, North Carolina State College

WILLIAM MCGEHEE[1]

A. INTRODUCTION

The purpose of this investigation is to study the verbal responses of elementary school children to a free word association test to determine if age or sex differences are revealed by differences in verbal responses An earlier paper (4*a*) dealing with the reaction times of the subjects, who also furnish the data for this study, gives in detail the method of conducting the experiment, the test used, and pertinent facts concerning the subject population It is, therefore, unnecessary to repeat these data except to indicate that subjects are 400 in number, 50 subjects at each age and sex level from seven through ten years.

The author of this study has encountered the same difficulty in handling the data which other investigators using free word association technique have encountered—that of evolving adequate categories for classifying verbal responses Logical and quasi-logical classifications are largely subjective, grammatical classifications seem of no value in the analysis of age-sex differences; and the inability of children at this age level to give accurate introspective reports mitigate against the use of pleasant-unpleasant-neutral categories. If verbal responses are treated primarily from a statistical view point, as did Wheat (9), there is a danger of placing too great reliance on the probability of recurrence of the same response in another like sampling and thus weakening any conclusions which might be reached.

The method finally evolved for treating the verbal responses does not claim to be superior in all respects to methods used in earlier studies. Indeed it has borrowed freely from these methods. It is believed, however, that the method of classification used has

---

*Received in the Editorial Office on October 1, 1937

[1]The writer acknowledges the valuable assistance and criticisms of Dr Paul L Boynton, George Peabody College, in the preparation of this study

minimized the errors involved in the use of logical or quasi-logical categories or in too great faith in the probable error of a percentage; and that, if age-sex differences are present in the data, these differences can be shown by the method of classification used in this investigation. The question is after all one of results obtained, or as Kent and Rosanoff (3, p. 28) in speaking of their system of classification said, "To what extent are the distinctions of this classification useful? An answer to this question can be found only in the results."

Therefore, the data have been analyzed from the following viewpoints: (*a*) community of response as indicated by preferred responses, (*b*) percentage of normal responses, (*c*) occurrence of phrase and clause responses, (*d*) occurrence of failure to react, (*e*) percentage of individual responses. Individual responses are classified as (*a*) non-logical reactions, (*b*) sound reactions, (*c*) repetition of the stimulus, (*d*) neologisms, (*e*) perseveration.

B. PREFERRED RESPONSES

The tendency of any one stimulus word to elicit a total number of different response words is limited. Kent and Rosanoff (3, p 8) found that the number of words which occurred in response to the stimulus words on their list varied from 280 for the word *anger* to 72 in response to *needle*. They say (3, p. 14), "the one tendency which appears to be almost universal among normal persons is the tendency to give in response to any stimulus word one or another of a small group of common reactions." This group, then, of common reactions would indicate closely the community of association in any given group in so far as words can be said to represent association of ideas.

In the case of children, as previously pointed out, Rosanoff and Rosanoff (7) found an increased tendency towards individual reactions in their responses when compared with the responses of adults This tendency decreased from earlier to later ages and the adult type of common respondence seemed fixed in the child by the time he reached eleven years of age. Otis (5) likewise found indications of chronological development in the type of responses, and substantiates the Rosanoffs in fixing eleven as the age where adult types of response become relatively fixed. Then, in so far as the community of response indicates a community of association, a progressive chronological development of this characteristic can be assumed from the above evidence.

The community of association in this investigation is studied from the point of view of preferred responses to each stimulus word used in the test Preferred responses, representing not just a single occurrence of a response but the recurrence of the same response several times, give a better indication of the community of association than do frequency tables. It is no guarantee, as Wheat (9) points out, that because a word appears once in a thousand responses that it will recur in another thousand responses. Nor, as Wheat failed to point out, is its occurrence ten, twenty, or thirty times a guarantee of its recurrence in the next similar sampling. Therefore, the problem of preferred responses has been approached from the viewpoint of comparison rather than the probability of recurrence, and any reaction which occurs with a frequency of six or more per cent of the total responses at any age-sex level to a particular stimulus word is classified as *preferred response* This classification of preferred responses is for comparative purposes only and for that purpose 6 per cent, 16 per cent, or 60 per cent could have been used as a standard without materially affecting the results as long as the standard was kept constant.

The most striking characteristic of the preferred responses found in this study is their apparent homogeneity for each sex and for all the ages in this experiment. The same word response occurs time and again to the same stimulus word at every age-sex level. Also, the highest ranking preferred response at any one age-sex level to a given stimulus usually appears as either the highest ranking response at the other age-sex levels or as second or third in rank. The most noticeable exceptions appear in the all-age responses as differences between the sexes and only appear then for the stimulus words *whip* and *tunnel* There is then very little difference between the kind of words which appear as responses to particular stimulus words, either for the two sexes or the different age levels, and would indicate a decided community of word associations among the groups in this study.

Differences between the sexes appear to some extent when the percentage of responses given by boys and girls to each stimulus word is examined. Among the stimulus words for which the preferred responses show a difference of five per cent or more between the sexes, boys react to 20 of the stimulus words with a higher percentage of preferred responses than do the girls Among the stimulus words where the difference is less than five per cent the boys again show a higher percentage of preferred responses than girls Evidently, while the preferred responses are usually the same words to particular stimulus words, the boys use more preferred responses than do the girls.

The data which are given in Table 1 are found when attention is turned to the total number of responses which may be classified as preferred responses for all the stimulus words The percentages in this table were obtained by adding the number of preferred responses occurring to each stimulus word in each age-sex group, and figuring the percentage on the basis of the total number of responses made to all the stimulus words. That is, 2,750 possible responses at each age-sex level and 5,500 possible responses for all ages of any one sex.

TABLE 1

PERCENTAGE OF PREFERRED RESPONSES—ALL STIMULUS WORDS

	Sex	
Age	B	G
7	51.24	48 66
8	50 65	50 66
9	56 58	51.86
10	55 64	49.13
All ages	53 53	50 38

The data in Table 1 tends to substantiate the hypothesis drawn in the preceding paragraphs that boys give a higher percentage of preferred responses than girls. At each age level except the eight-year level boys give a higher percentage than do girls of all ages. It must be noted, however, that the differences are very small especially at the seven- and eight-year levels, and that the smallness of these differences preclude any sweeping generalization as to sex differences in reacting with preferred responses. If any noticeable sex differences exist they are at the nine- and ten-year level, suggesting that community of association as represented by the occurrence of preferred responses to a free word association test is more homogeneous for the sexes at the seven- and eight-year levels than at the nine- and ten-year levels.

From the chronological viewpoint the tendency to react with preferred responses is more noticeable in boys nine and ten than in boys seven and eight. There is, however, little difference in this tendency in girls at any of the age levels investigated in this experiment. There is a slight increase in the percentage of preferred responses at the age levels of nine and ten over the age levels seven and eight when the responses of the subjects in this experiment are considered without regard to sex This can be seen from the data in Table 2.

TABLE 2

PERCENTAGE OF PREFERRED RESPONSES AT THE AGES OF SEVEN, EIGHT, NINE, TEN

Sex	B and G	B and G	B and G	B and G
%	49 95	50 66	54.22	52 39
Age	7	8	9	10

C PHRASES, CLAUSES AND SENTENCES, OR MULTIPLE-WORD RESPONSES

The frequency with which phrases, clauses and even sentences occur in the responses of children to free word associations has been noticed by all investigators who have used this technique with children as subjects. In fact, the frequency of such responses has been one point of differentiation between the adult's and child's response. It has been pointed out by Rosanoft and Rosanoff (7) that the tendency to respond with more than one word decreased with the increase in chronological age and Otis (5) has substantiated this hypothesis McElvee (4) and other investigators of the difference between responses of subnormal and normal children also have discovered that the tendency is more prevalent in subnormal than in normal children Evidently a decrease should be expected in the number of phrase, clause, and sentence responses with an increase in the chronological and mental ages of our subjects

The second fact to be emphasized is that responses in phrases, clauses, and sentences are a normal tendency in many children if not all, and may, as Otis (5, p 273) believes, be "a stage in language development where the sentence is the unit and represents a thought idea." As a matter of fact all the phrase, clause, sentence responses occurring in these data gave logical ideas in connection with the stimulus words with which they occurred, and the majority of them could be analyzed into an idea which could be expressed in one word. There was a tendency, not readily presented statistically, on part of some of the subjects to impress the investigator by explaining the meaning and significance of the stimulus word in their sentence responses. As Jung (1, p. 104) suggests, this may indicate "an inner want of satisfaction and a voidness of feeling." The data in this investigation can neither affirm nor disprove this suggestion. It is believed, however, that the occurrence of the phrase, clause, sentence response was not entirely caused by the child's inability to follow directions. This investigator found, as did Otis (5, p. 273), that "the instructions are carried out until a word is given that suggests no single word but a whole thought," and then there would be, in spite of earlier instructions, a response in form of a phrase, a clause, or a sentence.

In classifying responses as belonging to this type such obvious word formation activities as the response "flagpole" to the word "flag" and other word combinations which usually occur together as "George Washington," "United States," and "Sunday School" have been omitted. Table 3 gives the frequency of occurrence in terms of per cent based on the total number of possible responses to all the stimulus words at each age-sex level.

TABLE 3

PERCENTAGE OF OCCURRENCE OF PHRASES, CLAUSES AND SENTENCES

Age	B	G	B and G
7	7.98	5.48	6.73
8	5.89	4.63	5.26
9	6.80	3.96	5.38
10	9.24	6.15	7.70
All ages	7.47	5.04	6.26

The tendencies shown in this table are very clear in regard to sex differences. At each age level the boys show a higher percentage of this type of response than do girls and the total percentage for boys of all ages is 2.43 greater than that for girls.

The age-level differences are not as marked. For girls there is found a steady decrease in percentage at each age level except the ten-year level. For boys, though, there is a decrease from the seven- to the eight-years level, and an increase at the nine-year level. At the ten-year level is found the largest percentage of phrases, clauses, and sentences which occurs in any age or sex group. Differences in per cent of multiple-word responses at the different age levels are more evident when the per cent of both boys and girls are combined at each age level. Table 3 shows that the combined percentage of boys and girls resembles closely the percentage of *multiple-word* responses of boys alone

Two hypotheses may be advanced to explain this evidence which differs from that found by other investigators There might be a possible developmental stage at 10 years of age in both boys and girls which is reflected in this tendency to respond in the multiple-word manner No data were found, however, to substantiate this supposition Secondly, the increase may be due to a selected sampling either of stimulus words or cases It was found that the stimulus which elicited multiple-word responses at the 10-year-old level likewise elicited similar responses at other levels It must be concluded that, while these data do not disprove the hypothesis advanced by other investigators, they fail to substantiate the hypothesis that the frequency of multiple-word responses decreases as the chronological age increases

D FAILURES TO REACT

Failures to react, as well as multiple-word responses, have been found by earlier investigators to be more characteristic of children and subnormal individuals than of normal adults Rosanoff (6, p. 552) found that occurrence of failure to react decreased with an increase of both the chronological age and mental age of his subjects Wheat (9) found that the failures to respond to many of his stimulus words exceeded in percentage the preferred responses to the same stimulus word. Woodrow and Powell (11) and Otis (5) also found that failure to respond was characteristic of children.

The occurrence of this phenomenon has been explained variously. Jung (2), as is his wont, explains it in terms of emotional inhibition. Both Wheat (9, p 38) and Woodrow and Powell (11) suggest that in experiments where the Kent-Rosanoff word list has been used, the etiological factor is unfamiliarity with the stimulus words. Wheat says, concerning the effect of the type of stimulus word, that different kinds of stimulus words get different percentages of common responses, uncommon responses, and failures of response. It has been the experience of the investigator with free word association in this study that each of these explanations is in part true and that there might be added a further explanation, namely, that the necessity of choice between several potential responses is also an etiological factor in the failures to react which occurred in this experiment.

The data in this study are in agreement with those of Rosanoff, just quoted, to the effect that the frequency of failure to react decreases as the age of the subject increases. Table 4 indicates that a slight decrease in percentage of failures to respond is apparent from seven to ten years.

TABLE 4

PERCENTAGE OF OCCURRENCE OF FAILURES TO REACT

Age	Sex B	G	B and G
7	3.6	5.89	4.75
8	3.19	3.41	3.30
9	2.22	2.98	2.605
10	1.39	2.98	2.17
All ages	2.63	3.68	3.115

From this table it can be seen that girls give a consistently higher percentage of failures to respond than do boys. The data suggest that this difference is an innate difference, although it would require further investigation to prove this contention. It is obviously not due to the type of stimulus words, for the stimulus words which show the highest percentage of failures to respond (*afraid, dance, cheat, prison, help, stealing*) elicit failures to respond from boys and girls in a ratio similar to the total differences in occurrence of failures to respond to all stimulus words. Further, only four stimulus words have failures to respond which are not divided between the sexes, and to only one of these, *bed*, is a failure to respond on the part of girls.

E. INDIVIDUAL RESPONSES AND NORMAL RESPONSES

Five types of verbal reactions have been grouped for treatment under the general heading of individual responses. They are: (*a*) non-logical, (*b*) sound-reactions, (*c*) repetition of the stimulus, (*d*) neologisms, (*e*) perseveration. These types are equivalent to those listed by Rosanoff (6) under the heading of *individual reactions* and by Wheat (9), with the exception of *repetition of stimulus* and the *occurrence of phrases* and *clauses*, which he classifies as failures to react, as *uncommon reactions*. It is in this class of responses

TABLE 5

PERCENTAGE OF INDIVIDUAL RESPONSES FOR EACH AGE AND SEX GROUP

Age	7			8			9			10			All ages		
Sex	B	G	B-G	B	G	B-G	B	G	B-G	B	G	B-G	B	G	B-G
Non-logical	1 78	2 03	1.91	3 37	2 94	3 16	1 37	1.09	1 23	2 33	2.10	2.22	2 22	2 04	2 13
Sound reactions	1 02	1 53	1.28	1.93	1 16	1 56	80	73	.77	1 02	62	82	1 19	1.09	1 14
Repetition of stimulus	1 35	1 56	1 46	80	38	59	51	1 67	1 09	.66	57	62	83	1 04	.94
Neologisms	.29	29	.29	1.24	11	68	18	38	28	29	07	18	05	02	03
Perseveration	22	33	27	40	15	28	15	07	11	15	15	15	023	018	02
Individual responses	4 66	5 74	5 20	7 74	4 74	6.24	3 01	3.94	3 48	4 45	3 51	3 98	4.31	4.21	4.26

that we can expect to find any pathological reactions or "complex" indicators that might occur Hull and Lugoff found in their study of the strength of "complex" signs that each of the five above types showed some strength. In fact, these investigators declared that the *repetition of the stimulus* was the most reliable of the nine "complex" indicators used in their experiment. Jung (2), Kent and Rosanoff (3) and others have noted the relative frequency of the phenomenon of perseveration and of the occurrence of neologisms in neurotic and psychopathic patients. However, it is not meant to imply that all reactions falling in these five types are pathological or signs of emotional inhibition. This is especially untrue of children's *responses, for, as Otis (5) points out, these types of responses* may be merely manifestations of certain stages in the language development of the child

Table 5 gives the percentage of occurrence of each type of response which is classified in this study as an *individual response*

A brief description of each of the types of individual responses is necessary to this study for the classification of normal responses depends to a large measure upon the definition of individual responses

Non-logical responses in this classification were arrived at by the process of elimination. No response that could be fitted into any *of the four other categories among individual responses was* classified as non-logical. The failure-to-react type of response and the multiple-word type, because of their very nature, were automatically eliminated. This narrows the problem down to placing the responses into either normal or non-logical classification. Any response which was obviously without logical connection to the stimulus word was designated as belonging to the non-logical type In this category are placed responsiveness of the "explosion" types of which Kent and Rosanoff (3, p 21) said, "the stimulus word acts . . . merely as a signal for discharge." In addition certain responses were given which were due to the misunderstanding of the stimulus word, such as the response *Christmas* to *prison*, and the response, *iron*, *copper*, and *steel* to the stimulus word *cold* These responses are classified as non-logical.

The nature of the type of response classified as a neologism is obvious from the meaning of the term.

Responses have been classed as perseverative which are either (*a*) repetition of the preceding stimulus words, (*b*) repetition of the responses to the preceding stimulus words, (*c*) associated with the preceding stimulus word, (*d*) associated to the preceding response. This classification follows closely the one used by Kent and Rosanoff (3, pp. 22-23).

The classification of verbal reactions as *normal responses* has been arrived at by the process of elimination. All responses which do not fall into the categories given in the preceding pages of this chapter constitute the responses classified as *normal*. These responses are reactions which show logical or grammatical connections with the stimulus words with which they occur. These responses belonging to the normal category may be defined as descriptive terms, synonyms, anonyms, contiguous expressions, causal relationships, ideas of coordinate, superordinate, and subordinate relationship, word compounding and syntatical changes.

Normal responses of both girls and boys regardless of age tend to increase in percentage as the chronological age increases. The percentage of normal responses at the ten-year level is smaller than that of the nine-year level, but is greater than at the other two levels. The differences at each level, however, are so small that a general conclusion that the number of normal responses increases with age from seven to ten years is not justified.

Individual tendencies of both boys and girls, as is to be expected, show a tendency to decrease as the age increases. The noticeable exception to this is at the eight-year level where occurs the greatest percentage of individual responses found in this study. It is pos-

TABLE 6

PERCENTAGE OF OCCURRENCE OF INDIVIDUAL AND NORMAL RESPONSES

Age	Classification	B	G	B-G
7	Normal	83.76	82.89	83.82
	Individual	4.86	5.74	5.20
8	Normal	83.18	87.22	85.20
	Individual	7.74	4.74	6.24
9	Normal	87.97	89.12	88.14
	Individual	3.01	3.94	3.43
10	Normal	84.92	83.36	86.25
	Individual	4.45	3.51	3.98
All ages	Normal	85.61	87.13	86.37
	Individual	4.31	4.21	4.26

sible that the distribution of cases at this level may account for the increase in individual responses.

Table 6 gives the percentage of occurrence of normal responses at all age-sex levels.

The seeming contradiction found in Table 6 that there is both an increase in normal responses and in individual responses at the same age level can be reconciled by reference to Tables 3 and 4. It seems that a decrease in *multiple-word* responses, and, sometimes, in *failures-to-respond* is accompanied by an increase in responses classified in this study as individual responses.

Girls, when considered regardless of age level differences, give 1.5 more normal responses than do boys of all ages and 0.1 per cent less individual responses. These differences are so small that they lead to a tentative conclusion that the differences in percentage of normal responses between the two sexes at the age levels of seven *through ten are not significant.*

This foregoing conclusion can be drawn at each age level also, except at the eight-year level where the difference in percentage of normal and individual responses between the sexes is the greatest. This exception probably is due to the distribution of cases at the eight-year age level.

F. SUMMARY

The system of treating verbal responses and the method of classification have been indicated. In order to determine the extent of the community of ideas as shown by the word association technique in boys and girls seven, eight, nine, and ten years old the preferred responses have been examined. These consist of any response occurring six or more times to any stimulus word at any age-sex level. The conclusion was drawn that the community of ideas is not only high between girls at the same age level and boys at the same age level, but also between girls and boys at the same age level and between all the children that make up the subjects of this investigation. Certain tendencies towards differences which in the main are slight may be summarized as follows:

1. *Boys give a slightly higher percentage of preferred responses* than do girls. The difference is 3.15 per cent.

2. There is a slight tendency for the number of preferred re*sponses to increase with the increase in age, when members of both*

sexes are considered as a unit. This tendency is not consistent when boys or girls are considered alone The most decided difference in percentage of preferred responses is between the seven- and nine-year levels.

Multiple-word responses are apparently normal methods of response for most children The following tendencies in this type of response are to be noted:

1. Boys give more multiple-word responses than girls.

2. No definite tendency exists between age levels in difference in number of this type of response.

Failures to respond may be attributed either to emotional inhibition, antipathy towards the test situation, or an unfamiliarity with the stimulus words This type of response shows the following tendency:

Girls fail to respond more than do boys and this tendency decreases in both sexes with an increase in age

The method of classifying normal and individual responses in this study was by a process of elimination. Individual responses consisted of non-logical, sound reactions, repetition of the stimulus, neologisms, and perseveration Normal responses are all responses not classified as individual responses, multiple-word responses, or failures to react The following tendencies are seen in the percentage of normal and individual responses in this experiment.

1. Girls give fewer individual and more normal responses than do boys.

2 The number of normal responses tends to increase as the age of the subjects increase for both sexes.

It must be noted that the above tendencies are tendencies only. The verbal responses of the different sexes and of the different age levels differ by such small and inconsistent percentages that no sweeping conclusions can be drawn. In fact, the differences between the age-sex levels are slighter than the differences between individuals at a given age-sex level

REFERENCES

1 JUNG, C G Collected Papers on Analytical Psychology (2nd Ed) London Balliere, Tendall and Cox, 1920

2 ——— Studies in Word-Association New York Dodd, 1919

3 KENT, GRACE H, & ROSANOFF, A. J A study of association in insanity *Amer J Insan*, 1910, 67, 5-126

4. McELVEE, EDNA W Association in normal and sub-normal adolescents. *Amer J Psychiat.*, 1931-32, **11**, 331-318.

4a McGEHEE, W. The free word association of elementary school children I. Reaction times. *J Genet Psychol*, 1937, **50**, 441-455

5 OTIS, MARGARET. A study of association in defectives. *J. Educ. Psychol*, 1915, **6**, 271-288

6 ROSANOFF, A J. Manual of Psychiatry. New York. Wiley, 1920.

7 ROSANOFF, ISABEL R., & ROSANOFF, A J A study of association in children *Psychol. Rev*, 1913, **20**, 43-89

8 TOLMAN, E C, & JOHNSON, I A note on association time and feeling *Amer J Psychol*, 1918, **24**, 187-195.

9. WHEAT, L B. Free Associations to Common Words Contrib Educ, No. 498, Teach Coll, Bur. Publ, Columbia Univ., 1931.

10 WHITE, M N., & POWELL, N The differential reaction time for pleasant and unpleasant words *Amer J. Psychol.*, 1936, **48**, 126-133.

11 WOODROW, H., & POWELL, F Children's association frequency tables *Psychol Monog*, 1916, **22**, No 5

12 WYATT, H G. Free word association and sex differences. *Amer J Psychol*, 1932, **44**, 454-472.

Department of Psychology
North Carolina State College
Raleigh, North Carolina

The Journal of Genetic Psychology, 1938, 52, 375-393

LATERALITY IN MONKEYS*

Child Welfare Research Station, State University of Iowa

J S. KOUNIN[1]

A INTRODUCTION

The topic of handedness has recently acquired added significance because of its relationship to speech pathology (13), to introversion-extraversion and intelligence (2); and to cortical functions concerned with hemispheral dominance (17). Most theories of the origin of handedness assume that preferential handedness is characteristic only of man. [A detailed treatment of this topic is given by Wile (16) and Ludwig (10)]. Partly because of definite social pressures operating to produce dextrality, the determination of the nature of handedness in human beings is difficult and complicated. In an effort to overcome this difficulty some investigators (11, 14) have used rats in studies of handedness These experiments have added considerably to our knowledge of handedness and have also pointed to the existence of preferential handedness among infrahuman animals. If preferential handedness could be demonstrated in monkeys, their structural and functional affinity to man would argue for their utilization in studies of handedness. Moreover, it would be possible to investigate the nature of general sidedness in monkeys as related to handedness and eyedness

The following study has been carried out mainly (*a*) to develop techniques for the study of handedness and eyedness in monkeys; (*b*) to determine whether monkeys possess preferential handedness.

B PERTINENT LITERATURE

There is insufficient experimental material dealing with handedness in subhuman primates. Primarily, whether they have preferen-

---

*Received in the Editorial Office on October 9, 1937.

[1] The author wishes to express his indebtedness and thanks to Professor Harry F. Harlow who encouraged and actively cooperated in this study which was conducted during the year of the author's appointment as Research Assistant at The University of Wisconsin Primate Laboratory. Further indebtedness and thanks are due Dr Paul Settlage and Mr Walter Grether for their many valuable suggestions and aids

tial handedness has not, as yet, been conclusively determined; in the second place, no techniques for testing handedness have been standardized. Cunningham (1) reviewed the literature prior to 1902 regarding the existence of handedness in monkeys, and found considerable disagreement among authorities. Moreover, most studies at that time were observational.

Since that time the following observational and experimental studies have been reported. After noting the hand used by six rhesus monkeys for picking up food, Franz (3) tentatively classified three as left-handed, one as right-handed, and two as ambidextrous. However, Sarasin (12) observed the hand three monkeys used for picking up food and concluded that monkeys are essentially ambidextrous. The observations of Yerkes (18) indicated the possibility that monkeys and apes manifest preferential handedness when reaching for food or when manipulating objects in connection with food getting. Lashley (9) found some agreement, in two rhesus monkeys, between the hand used more often for reaching food and the hand used more frequently in warding off the touch of the experimenter. With the exception of Yerkes' observations of tool-using behavior in two apes, these studies have regarded the hand used in over a chance majority of the time for picking up or reaching for food as the criterion of handedness in monkeys.

A pertinent observation was made by Klüver (7) with respect to the problem of testing techniques. He noted that, although a preferential handedness was not manifested for pulling in food boxes when the end of the string was placed in the monkey's cage, it was expressed when the end of the string was placed at some distance from the cage. Similarly, unpublished studies of Harlow, Settlage, and Grether (5) at the University of Wisconsin Primate Laboratory indicated that the degree of hand-finger manipulation involved must be considered when testing for handedness in monkeys.

C. Animals, Housing and Care

The monkeys used in this study were housed and cared for at the University of Wisconsin Primate Laboratory. Table 1 gives a brief description of each monkey.

TABLE 1

Monkey	Classification	Sex	Age
Duke	Rhesus	Male	Pre-Pubescent
Peggy	Rhesus	Female	Pubescent
Jyppo	Rhesus	Male	Pre-Pubescent
Jake	Rhesus	Male	Pre-Pubescent
Johnny	Spider	Female	Pre-Pubescent
Murphy	Cebus	Male	Pubescent
Dusty	Cebus	Male	Pubescent
Slit-Ear	Cebus	Male	Pubescent

D APPARATUS AND PROCEDURE

Three tests of handedness and one of eyedness were used in this study, and 100 trials were obtained for each test, according to the following procedure Testing was started at the beginning of a week and 25 trials made on each of 2 succeeding days A rest day was then interposed and 25 trials per day again obtained on the 2 days following The eyedness test data were secured during the first handedness test week, and the two remaining handedness tests were run on the two weeks following At least a month after the last trial of the first test week, the entire experimental series was repeated Thus the results include 200 trials for each test situation.

Preliminary trials, in random sequence, were made with six of the monkeys with the first handedness test to accustom them to the experimental cage and to the experimenter.

A description of the tests of handedness and eyedness follows

1. *Handedness* I.

The monkeys were tested in a circular sheet metal cage 37 in in diameter and 30 in high. This cage rested upon a table in the middle of which was a hole 2 in. in diameter A food container 1 in in diameter and ¼ in. deep, which could be raised or lowered, was placed directly beneath this hole The distance from the floor to the container varied from 8 to 11 in, depending upon the arm length of the monkey To obtain a piece of banana, the monkey needed to extend his arm through the hole in the floor and remove the food from the container which was placed at a distance approximating his maximum reach

2. *Handedness* II.

The test cage, made of heavy welded 2-in. wire mesh, was 26 in high, 18 in. deep, and 28 in. wide. To the front center of the cage was fastened a thin piece of ply wood 9 in. high and 7 in wide. A hole 2 in. in diameter was drilled in this board 3 in. from the bottom of the cage. A food box 2 in square, with a spring cover 2¼ in. square, was centered and securely attached at a distance 7 to 9 in. from the front of the cage, depending upon the monkey's reach. To obtain a piece of banana the monkey needed to put one hand through the hole in the ply wood, lift the cover of the food box, hold it open, and remove the reward from a narrow slanting depression in the food box—only one hand could be used for the entire manipulation

3. *Handedness* III.

The Handedness II cage was used here, the ply wood removed, and the front covered with ¼ in. hardware cloth. An opening 8 in high and 3 in. wide was made in the front center of the cage, to permit the free use of either hand. To obtain food, which was beyond reach, the monkey was required to manipulate a rake. The handle of the rake was 10 in. long, 1 in. in diameter, and the cross bar 3 in. wide. The nearest end of the rake was placed from 8 to 10 in. in front of the opening. The food was alternately placed on opposite sides of the rake handle directly ahead of the opening of the cage. The hand used for pulling in the rake was recorded in each trial.

4. *Eyedness Test.*

A hole ½ in. in diameter was drilled in front of the aforementioned sheet metal cage and a straight porcelain tube with a ⅛ in. opening fitted into it This tube projected ½ in into the cage, one foot from the floor. Approximately 5 in. above this a small door was constructed which could be opened and closed by the experimenter. Whenever a monkey sighted through the tube, in response to a slight noise and food placed at a short distance in front of the tube, the door was opened and the animal was rewarded with a piece of banana The eye used for looking through the tube was observed by means of a mirror placed above an opening in the top of the cage. For the retest series, after a month, a tin V-scope slightly over 2 in long was constructed in the door, with the narrow end extending ½ in. inside the cage. The narrow end was ½ in. in diameter and the wide end 2 in in diameter. As well as to

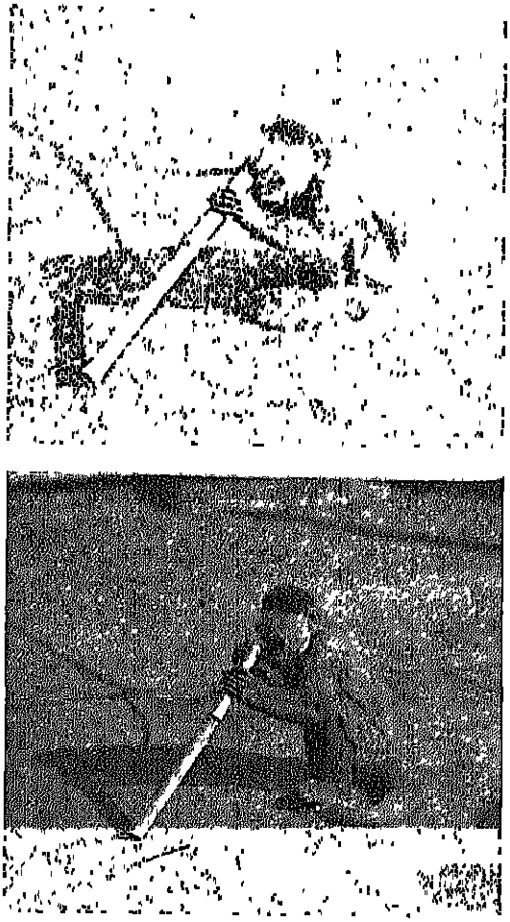


FIGURE 1

PHOTOGRAPHS DEMONSTRATING EYEDNESS IN MURPHY, A CEBUS MONKEY
Although they do not illustrate the apparatus used in this experiment, they reveal the same general principle

increase the visual range of the monkey, this device was used to check the original test Since the height was changed, the monkeys had to assume a different position, thereby reducing the influence of any possible position habit developed in the first series upon the second series

TABLE 2

THE NUMBER OF TIMES EACH HAND WAS USED BY FOUR RHESUS MONKEYS FOR OBTAINING FOOD IN THE DESIGNATED EXPERIMENTAL SITUATIONS
The eye used in sighting is indicated for one monkey.

	Handedness I			Eyedness		Handedness II			Handedness III		
		Hand		Eye			Hand			Hand	
Monkey	Date	R	L	R	L	Date	R	L	Date	R	L
Duke	1/8-1/28	3	58								
	3/25	0	25			3/3	19	6	3/16	2	23
	3/26	0	25			3/4	6	19	3/17	0	25
	3/28	0	25			3/6	2	23	3/19	4	21
	3/29	0	25			3/7	1	24	3/22	5	20
	4/26	0	25			4/7	0	25	4/26	17	8
	4/27	0	25			4/8	0	25	4/27	6	19
	4/29	0	25			4/10	0	25	4/29	6	19
	4/30	0	25			4/11	0	25	4/30	0	25
	Per cent	1	99				14	86		20	80
Peggy	12/2-1/11	315	29								
	3/25	25	0	0	25	3/3	25	0	3/16	20	5
	3/26	25	0	0	25	3/4	25	0	3/17	16	9
	3/28	25	0	0	25	3/6	25	0	3/19	14	11
	3/29	25	0	1	24	3/7	25	0	3/22	21	4
	4/26	25	0	0	25	4/7	25	0	4/26	1	24
	4/27	25	0	0	25	4/8	25	0	4/27	0	25
	4/29	25	0	0	25	4/10	25	0	4/29	0	25
	4/30	25	0	0	25	4/11	25	0	4/30	1	24
	Per cent	94	6	.5	99.5		100	0		37	63
Jyppo	3/25	14	11			3/3	24	1	3/16	22	3
	3/26	10	15			3/4	25	0	3/17	14	11
	3/27	0	25			3/5	25	0	3/19	21	4
	3/28	1	24			3/7	25	0	3/22	25	0
	4/3	11	14			4/7	25	0			
	4/4	0	25			4/8	25	0			
	4/10					4/25	25	0			
						4/11	25	0			
	Per cent	22	78				99.5	5		84	16
Jake						3/3	13	12			
						3/4	11	14			
						3/6	3	22			
						3/7	0	25			
						Per cent	27	73			

TABLE 3

The Number of Times Each Hand and Eye Were Used by Three Cebus and One Spider Monkey in the Designated Experimental Situations

Monkey	Handedness I Hand Date	R	L	Eyedness Eye R	L	Handedness II Hand Date	R	L	Handedness III Hand Date	R	L
Murphy	11/17-1/9	432	1								
	3/2	25	0	25	0	2/23	24	1	3/11	25	0
	3/3	25	0	25	0	2/24	24	1	3/18	25	0
	3/5	25	0	25	0	2/26	25	0	3/19	25	0
	3/6	25	0	25	0	2/27	25	0	3/22	25	0
	4/7	25	0	25	0	3/29	24	1	4/26	25	0
	4/8	25	0	25	0	3/30	25	0	4/27	25	0
	4/10	25	0	25	0	4/3	25	0	4/29	22	3
	4/13	25	0	25	0	4/4	25	0	4/30	16	9
	Per cent	99 8	2	100	0		98 5	1 5		94	6
Dusty	11/20-1/8	25	0								
	3/2	25	0	25	0	2/23	25	0	3/11	25	0
	3/3	25	0	25	0	2/24	25	0	3/18	25	0
	3/5	25	0	25	0	2/26	25	0	3/19	25	0
	3/6	25	0	25	0	2/27	25	0	3/22	25	0
	4/7	25	0	25	0	3/29	25	0	4/26	25	0
	4/8	25	0	25	0	3/30	25	0	4/27	25	0
	4/10	25	0	25	0	4/3	25	0	4/29	25	0
	4/13	25	0	25	0	4/4	25	0	4/30	25	0
	Per cent	100	0	100	0		100	0		100	0
Slit-Ear	11/20-1/18	8	228								
	3/2	0	25	25	0	2/23	0	25	3/11	1	24
	3/3	0	25	25	0	2/24	0	25	3/18	3	22
	3/5	0	25	25	0	2/26	0	25	3/22	9	16
	3/6	0	25	25	0	2/27	0	25	3/23	7	18
	4/7	0	25	25	0	3/29	0	25	4/26	0	25
	4/8	0	25	25	0	3/30	0	25	4/27	2	23
	4/10	0	25	25	0	4/3	0	25	4/29	0	25
	4/13	1	24	25	0	4/4	0	25	4/30	1	24
	Per cent	22	98	100	0		0	100		11.5	88.5
Johnny	11/23-1/8	105	6								
	3/2	24	1			2/23	0	20	3/11	2	23
	3/3	25	0			2/24	0	25	3/18	13	12
	3/5	25	0			2/26	0	25	3/19	24	1
	3/6	25	0			2/27	0	25	3/22	25	0
	4/7	25	0			3/29	0	25	4/26	25	0
	4/8	25	0			3/30	0	25	4/27	25	0
	4/10	25	0			4/3	0	25	4/29	25	0
	4/13	25	0			4/4	0	25	4/30	25	0
	Per cent	98	2				0	100		82	18

E. Results

The daily results for each animal on the different tests are given in Tables 2 and 3. These results will be analyzed in terms of various criteria of consistency, differences between the tests used, handedness of the individual monkeys, and differences between the species of monkeys tested.

1. *The Criteria.*

The criteria of handedness used in summarizing the results of each test were selected arbitrarily They are:

1 One hand used in at least 90 per cent of the total trials,

2. One hand used in at least 85 per cent of the total trials;

3 One hand used in at least 80 per cent of the total trials,

4 One hand used in at least 75 per cent of the trials on *every day* the animal was tested.

5. No shifts in handedness at any time. A shift is defined as the use of the least preferred hand (as determined up to that time in any single test series) in 51 per cent or more of the trials on any *one day*. Hence, a maximum of four shifts is possible in the eight test days of any one test situation.

TABLE 4

Summary Table Showing the Criteria Seven Monkeys Failed to Meet in the Designated Tests, the Total Number Failed by Each Monkey, and the Total Number Failed on Each Test*

	Handedness I Criterion					Handedness II Criterion					Handedness III Criterion					Total number failed by each monkey
Monkey	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	
Duke						F				1	F	F		1	1	6
Peggy											F	F	F	1	4	5
Jyppo	F	F	F	3	1						F	F		1		8
Johnny											F	F		1	1	4
Murphy														1		1
Dusty																0
Slit-Ear											F				2	2
Total number failed			5					2					19			19

*The totals refer to the number of criteria failed, not to the absolute number of days of failure, that is, any single criterion failed on several days is treated as one failure

The results of this analysis are summarized in Table 4. The criteria are numbered to correspond with the description above An "*F*" indicates which of the first three each monkey failed to meet on the various tests. The numbers under Criterion 4 refer to the number of days this condition was not satisfied. The numbers under Criterion 5 designate the number of days a "shift" occurred The data reveal that the degree of preference evidenced is significantly higher in Handedness I and II than in Handedness III, and that all three cebus monkeys failed fewer criteria than any of the others

2 *Intra-test Consistency*

Intra-test consistency refers to the retest reliability of each separate test. As noted in the "Methods," each test was given in two series of 100 trials each, the second at least one month after the last previous trial of the first series. The reliability of each test can, therefore, be determined by comparing the relative degree of hand preference expressed in the two series Table 5 summarizes the intra-test consistency A high per cent difference indicates a low reliability This analysis indicates that Handedness III is considerably less reliable than any of the other tests Since the relative difference between Handedness I and Handedness II devolves almost entirely on the inconsistency of a single monkey, this difference is not considered significant

TABLE 5

SUMMARY TABLE SHOWING THE DIFFERENCE IN PER CENT FREQUENCY OF HAND PREFERENCE BETWEEN THE FIRST AND SECOND SERIES OF THE DESIGNATED TESTS

Monkey	Handedness I	Handedness II	Handedness III	Eyedness
Duke	0%	28%	18%	No results
Peggy	0%	0%	69%	1%
Johnny	1%	0%	36%	No results
Dusty	0%	0%	0%	0%
Murphy	0%	1%	12%	0%
Slit-Ear	1%	0%	17%	0%
Total difference	2%	29%	152%	1%

3. *Inter-test Consistency.*

Inter-test consistency is concerned with the degree of agreement

TABLE 6

SUMMARY TABLE SHOWING THE DEGREE OF HAND PREFERENCE MANIFESTED BY SEVEN MONKEYS IN THE FIRST AND SECOND SERIES OF THE DESIGNATED TEST

Monkey	Handedness I Preliminary series	Handedness I First series	Handedness I Second series	Handedness II First series	Handedness II Second series	Handedness III First series	Handedness III Second series
Duke	95 % L	100% L	100% L	72% L	100% L	89% L	71% L
Peggy	91 % R	100% R	100% R	100% R	100% R	71% R	98% L
Jyppo		78% L		99% R	100% R	84% R	
Johnny	95 % R	99% R	100% R	100% L	100% L	64% R	100% R
Murphy	99 8% R	100% R	100% R	99% R	98% R	100% R	88% R
Dusty	100 % R	100% R	100% R	100% R	100% R	100% R	100% R
Slit-Ear	97 % L	100% L	100% L	100% L	100% L	80% L	97% L
Mean percentage preference		97 6%		97.7%		87.8%	

between the different tests of handedness Before the existence of preferential handedness is established in any monkey, it is necessary that it manifest a preference for the same hand on all three tests. The amount of inter-test consistency is indicated in Table 6 This table shows both that a higher per cent of preference is expressed on Handedness I and Handedness II, and that three of the monkeys show no consistent preference on the different tests used

4 *Differences Between the Tests.*

It is evident that the results are dependent upon the test used The greatest degree of hand preference is expressed in Handedness I and II Table 4 shows that the least number of criteria are failed on these two tests, Table 6 indicates further that the highest per cent of hand preference is expressed on these two tests; and, in addition, these two tests possess the greatest intra-test consistency (reliability), as revealed in Table 5. Handedness III, however, possesses the least diagnostic value in differentiating hand preference A large number of criteria are not met in this test (Table 4), it shows the lowest per cent of hand preference (Table 6), and it has the least reliability (Table 5).

5 *Handedness of the Individual Monkeys*

Duke manifested a definite and consistent preference for the left hand on Handedness I according to all five criteria. He used his left hand in 86 per cent of the total trials on Handedness II and shifted to the use of the right hand on one of the eight days tested, thereby failing Criterion 1 and Criterion 5 for this test. The only standard he did attain on Handedness III was the use of the left hand in at least 80 per cent of the total trials. Generally, then, Duke possessed a moderate, but slightly inconsistent, left hand preference.

Peggy showed a decided preference for the right hand on both Handedness I and Handedness II according to all five criteria, but failed to support this choice on Handedness III, completely reversing to a left hand preference in the second series

Although Johnny met all criteria on Handedness I and Handedness II, he used the left hand in one test and the right hand in the other, indicating ambidexterity.

The limited data for Jake (on Handedness II only, a reliable test) suggests no significant preference.

Murphy and Dusty can be said to possess definite handedness They used the right hand almost exclusively on all three tests. (There is no history of injury to Dusty's left hand, while Murphy's use of the left hand for nine of the trials on Handedness III can probably be attributed to an injury to his right index finger which was bitten by another monkey the previous day)

Slit-Ear is *consistently left-handed on Handness I and* Handedness II but fails to satisfy the two most rigid criteria on Handedness III. However, since the latter test has been found to express the least degree of hand preference, Slit-Ear's use of the left hand in 88.5 per cent of the total trials on this test is highly significant. The data on this animal argues for the justifiability of attributing preferential handedness to it.

6. *Differences Between the Species of Monkeys Tested.*

The data show noticeable differences between the cebus monkeys (Murphy, Dusty, and Slit-Ear) and the others tested. The cebus monkeys undoubtedly possess a greater degree of preferential handedness than do any of the others. None of the cebus monkeys failed any of the criteria for Handedness I and II, as indicated in Table 4. Johnny and Peggy are the only other monkeys tested who failed no criteria on these two tests, but Johnny used a different hand for each test, while Peggy reversed hands on Handedness III. Species' differences are especially noticeable on Handedness III. All three cebus monkeys show a greater per cent of preference on this test than do any of the others (Table 6); express greater consistency between the first and second series (Table 5), and fail fewer—and only the most severe—criteria (Table 4). As indicated in Section 5, the cebus monkeys are the only ones who were definitely handed, according to the criteria employed.

7. *Eyedness.*

Preferential eyedness, as measured by the method used, was definite and consistent for the monkeys tested. Attempts to obtain data from the other monkeys were unsuccessful. Table 7 shows that eyedness does *not necessarily correspond with handedness.*

F. DISCUSSION OF RESULTS

The finding that cebus monkeys (and possibly some of the others, if less rigid criteria are employed) possess preferential handedness is

TABLE 7

SUMMARY TABLE SHOWING THE EYEDNESS AND HANDEDNESS OF FOUR MONKEYS, EXPRESSED AS PER CENT PREFERENCE IN DESIGNATED SITUATIONS

Monkey	Eyedness R	Eyedness L	Handedness I R	Handedness I L	Handedness II R	Handedness II L	Handedness III R	Handedness III L
Murphy	100 %	0 %	99.8%	2%	98.5%	1.5%	94 %	6 %
Dusty	100 %	0 %	100 %	0 %	100 %	0 %	100 %	0 %
Slit-Ear	100 %	0 %	2 %	98 %	0 %	100 %	11.5%	88.5%
Peggy	.5%	99.5%	94 %	6 %	100 %	0 %	37 %	63 %

contradictory to the opinion of various authorities, including Sarasin (12), Ludwig (10), and Cunningham (1). As noted in the introduction, however, the hand used for picking up food was the only experimental measure of handedness. It is doubtful whether a hand preference, if present, can be reliably measured in the execution of such a task, since unnoticeable posturings and situational expediencies play too important a part. [It is worth noting that Updegraft (15), in a study of preferential handedness in young children, excluded the "hand used for picking up five small articles" and the "hand used for picking up a spoon from the floor" from a battery of handedness tests because of the low reliability of these two measures.] Hand usage is no simple all-or-none affair. Tests for handedness should include performances that involve some degree of delicate and difficult manual manipulation. It is highly probable that the existence of handedness in man could be "disproved" by using one or two inadequate tests.

Studies of handedness in man also demonstrate that the method of measuring hand preference influences greatly the nature of the results obtained. In a thorough study of handedness in man, Koch (8) has shown that tests vary in their effectiveness in determining handedness. Tasks which she calls "holding and carrying activities" possess relatively low diagnostic value. In activities requiring little skill she found a tendency towards indiscriminate use of either hand. Tests which involved more refined and elaborate manipulation possessed the greatest handedness-index value.

The results of Harlow, Settlage, and Grether (5) indicate that the most effective handedness tests were those demanding a high degree of coordinated manual skill. They tested eight monkeys in three situations; 25 trials were employed in a single test period, and each test was given three times in the first series. The entire procedure was repeated after a period of approximately five weeks, giving a total of 150 trials for each test. A brief description of their tests follows: (*a*) In the pipe test two short pipes were placed through a board which was hung against the front of the monkey's cage Food was placed in front of each pipe, in irregular order, and food was retrieved by the monkey through the respective pipe. (*b*) The cup test consisted of a shallow food container which was placed approximately 6 in. in front of the cage and from which the monkey was to remove a piece of food (*c*) To secure food in the string test, the monkey pulled in a string which was approximately 3 in. from the front of the cage Their results are summarized in the following three tables (The pipe test is treated as two separate tests in the tables. An equal number of trials was used both for the pipe on the right side of the board and for the pipe on the left side.)

Their results disclose the greatest degree of hand preference in the pipe test and the least in the string test. The pipe test reveals the highest average per cent of hand preference (Table 8), the greatest reliability (Table 9); and the least number of criteria failed (Table 10). In contrast, the string test shows the lowest average per cent of hand preference, the least reliability, and the most criteria failed. The cup test is intermediate between the two other tests according to all the above indexes

TABLE 8

THE PER CENT FREQUENCY OF THE HAND PREFERRED BY EIGHT MONKEYS IN 150 TRIALS IN THE DESIGNATED TESTS

From data of Harlow, Settlage, and Grether (5).

Monkey	Pipe on left	Pipe on right	Cup	String
Java 1	66% R	64% R	91% L	76% R
Java 2	100% L	100% L	96% L	100% L
Rhesus 1	93% R	95% R	100% R	63% R
Rhesus 2	98% R	64% R	63% R	94% R
Rhesus 3	93% L	95% L	55% L	73% R
Rhesus 4	97% R	90% R	69% R	57% L
Rhesus 5	98% L	98% L	51% R	61% R
Pigtail	65% L	94% L	83% L	63% L
Mean per cent preference	88.5%	87.7%	76%	73.3%

If the same criteria of handedness used in this study are applied to the interpretation of their results, then Java 2 was the only de-

TABLE 9

SUMMARY TABLE SHOWING THE DIFFERENCES IN PER CENT FREQUENCY OF HAND PREFERENCE BETWEEN THE FIRST AND SECOND SERIES OF THE DESIGNATED TESTS

From data of Harlow, Settlage, and Grether (5).

Monkey	Pipe on left	Pipe on right	Cup	String
Java 1	18%	6%	5%	33%
Java 2	0%	0%	3%	0%
Rhesus 1	11%	7%	0%	17%
Rhesus 2	1%	47%	36%	12%
Rhesus 3	4%	6%	72%	53%
Rhesus 5	1%	1%	19%	31%
Rhesus 5	1%	1%	19%	31%
Rhesus 4	7%	17%	13%	28%
Pigtail	11%	0%	19%	13%
Total	53%	84%	167%	187%

finitely handed monkey. However, it should be noted that the criteria suggested in this paper are considerably more rigid than in the usual studies of handedness in man and rats. (Such standards were employed because of the prevailing opinion among authorities that monkeys are ambidextrous) The results suggest, that for further refinement of the testing technique the criteria be determined by the nature of the test. The more highly diagnostic the test (as Handedness I and Handedness II), the more rigid the criteria The summarized results of Harlow, Settlage, and Grether (5) in Table 8 indicate a decided handedness in Java 2 and some degree of handedness in Rhesus 1, Rhesus 2, and the pigtail monkey The other four monkeys did not express any significant hand preferences. When due consideration is given to the relative value of the various tests used in both these studies, it becomes apparent that there is no more reason for concluding that rhesus monkeys, as a whole, are ambidextrous than for concluding that they are handed

The results of both experiments strongly suggest the advisability of speaking rather of degree than of a unit characteristic of handedness in monkeys. [Such an interpretation is also applied to handedness of man by Koch (8), Downey (2), Travis (13) and others.] This degree of handedness is greatly determined by the nature of the tests used Such tasks as Handedness I and Handedness II reveal considerably higher degrees of handedness than the cup test, string test, pipe test, and Handedness III. Analysis of the different test

TABLE 10

SUMMARY TABLE ANALYZING THE DATA OF HARLOW, SETTLAGE, AND GRETHER (5) IN TERMS OF THE CRITERIA OF SECTION E2 OF THIS PAPER, SHOWING THE CRITERIA EIGHT MONKEYS FAILED TO MEET IN THE DESIGNATED TESTS, THE TOTAL NUMBER FAILED BY EACH MONKEY, AND THE TOTAL NUMBER FAILED ON EACH TEST

Monkey	Pipe on left Criterion					Pipe on right Criterion					Cup Criterion					String Criterion					Total failed
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	
Java 1	F	F	F	2	2	F	F	F	2	1						F	F	F	1		14
Java 2																					0
Rhesus 1					1											F	F	F	1	4	6
Rhesus 2						F	F	F	2	1	F	F	F	3	2					1	11
Rhesus 3											F	F	F		1	F	F	F	1	2	9
Rhesus 4										1	F	F	F	2	3	F	F	F	2	2	11
Rhesus 5											F	F	F	2	4	F	F	F	1	3	10
Pigtail	F	F	F	2	4						F	F			2	F	F	F	2	4	13
Total number failed			11					11					22					30			74

situations reveals that those which manifest the greatest degree of hand preference demand more difficult, refined, or complex manipulations

It appears advisable to use a relatively unskilled performance, as Handedness III, in addition to several discriminative manipulations, when testing for handedness in monkeys While Handedness I and II possess the greatest diagnostic value, Handedness III, or a similar test, is more apt to reveal the degree of handedness It is this latter test which most clearly differentiates between the cebus and the other monkeys tested Thus, while Peggy and Duke manifested considerable handedness on Handedness I and II they failed to express a definite handedness on Handedness III. On the other hand, since the cebus monkeys prefer one hand consistently on all three tests, it is obvious that they are handed to a greater degree

An explanation of why a considerably more pronounced hand preference is found in the cebus than in the rhesus monkeys might be sought in the behavioral differences between the two species. Limited observations indicate a greater degree of general manipulatory and investigatory behavior among the cebus monkeys Those used in this study frequently reached and searched in the experimenter's pockets The experimenter has never noted this behavior in the other monkeys True tool-using has been noted in cebus but not in rhesus monkeys Klüver (7) has also found tool-using among cebus monkeys. Eyedness trials were obtained without difficulty in the cebus monkeys and they often looked through the tube used for the eyedness test before the experiment was started. Even after considerable training it was possible to obtain eyedness data for only one of the other monkeys It is hypothesized that the more clearly defined handedness of the cebus monkeys is closely related to, if not actually dependent upon, their generally increased manipulative behavior

There is evidence to indicate a progressively increasing degree of handedness during human infancy, and possibly, though less conclusively up to six years of age Jones (6) further points out that hand preferences are not reported to begin until the second half of the first year While maturation and learning undoubtedly are important factors, the rôle of increased manipulative behavior should also be considered. [According to the norms of Gesell and Thompson (4), the power of prehension is lacking in infants during the first three months, becomes more unilateral and refined in the next six months, and attains patterns closely resembling those of adults in the last three months of the first year. The amount of manipulation of environmental objects also increases with age] In line with the results obtained in this study, the increased prehension and manipulation of infants with increase in age, can be related to their increased handedness, both because it permits the use of more adequate and discriminative tests of handedness, and because it points to a greater degree of handedness being associated with increased manipulative behavior.

Studies with human beings also indicate some disagreement between handedness and eyedness. The fact that the eyedness of some monkeys does not correspond with their handedness makes it difficult to explain their eyedness solely in terms of hemispheral dominance Moreover, it is generally agreed that each eye is bilaterally represented in the cortex.

G. Summary and Conclusions

Four rhesus, one spider, and three cebus monkeys were studied with three different tests of handedness and one of eyedness. With regard to techniques of testing handedness, the results indicated the necessity of using more than one test, including different types of manipulation, and extending the tests over a reasonable length of time. The degree of preferential handedness manifested was largely dependent upon the nature of the task performed, the most effective handedness tests being those which required a greater degree of difficult, coordinated, and delicate manipulations Relatively unskilled performances revealed less marked and consistent handedness.

Eyedness data were obtained from only four of the monkeys. Their eyedness was definite and consistent, and did not necessarily correspond to their handedness.

A marked degree of preferential handedness was expressed by all three cebus monkeys. The other monkeys evidenced, as a whole, neither a decided handedness nor an essential ambidexterity. The considerably more pronounced handedness of the cebus monkeys was tentatively related to their generally increased investigatory and manipulative behavior

REFERENCES

1. CUNNINGHAM, D J. Right-handedness and left-handedness. *J Anthrop Inst*, 1902, **32**, 278
2. DOWNEY, JUNE E. Types of dextrality and their implications *Amer J Psychol*, 1927, **38**, 317-367
3. FRANZ, S. I Observations on the preferential use of the right and left hands by monkeys. *J Anim. Behav*, 1913, **3**, 140-144
4. GESELL, A., & THOMPSON, H. Infant Behavior, Its Genesis and Growth New York, McGraw-Hill, 1934 Pp viii+343
5. HARLOW, H F., SETTLAGE, P, & GRETHER, W Handedness in monkeys Unpublished studies at the University of Wisconsin Primate Laboratory, 1933
6. JONES, H E Dextrality as a function of age. *J Exper Psychol*, 1931, **14**, 125-143.
7. KLUVER, H Behavior Mechanisms in Monkeys. Chicago Univ Chicago Press, 1933 Pp xvii+387.
8. KOCH, H L., *et al* A study of the nature, measurement, and determination of hand preference. *Genet. Psychol Monog.*, 1933, **13**, 117-221
9. LASHLEY, K. S. Modifiability of the preferential use of the hands in the rhesus monkey. *J Anim Behav.*, 1917, **7**, 178-186
10. LUDWIG, W. Das Rechts-Links Problem ins Tierreich und beim Menschen Berlin. Springer, 1932. Pp. xi+496
11. PETERSON, G M Mechanisms of handedness in the rat *Comp Psychol Monog*, 1934, **9**, 1-67
12. SARASIN, P Über Rechts und Linkshandigkeit in der Prahistorie und Rechtshandigkeit in der Historischen Zeit *Verh Naturf Ges. Basel*, 1918, **29**, 122-196.
13. TRAVIS, L E Speech Pathology New York· Appleton, 1931. Pp xxxiv+331.
14. TSAI, L S, & MAURER, S Right-handedness in white rats. *Science*, 1930, **122**, 436-438.
15. UPDEGRAFF, RUTH Preferential handedness in young children. *J Exper. Educ*, 1932, **1**, 134-139
16. WILE, IRA S. Handedness, Right and Left Boston Lothrop, Lee and Shepard, 1934 Pp xiii+439
17. WEISENBURG, T M., & MCBRIDE, K E Aphasia· A Clinical and Psychological Study New York Commonwealth Fund Pp xvi+634
18. YERKES, R M The mental life of monkeys and apes· A study of ideational behavior *Behav. Monog*, 1916, **3**, 1-145.

Child Welfare Research Station
State University of Iowa
Iowa City, Iowa

The Journal of Genetic Psychology, 1938, 52, 395-400

PERFORMANCES OF SCHOOL CHILDREN ON THE REVISED STANFORD-BINET AND THE KENT E-G-Y-TEST*

New York Hospital and the Department of Psychiatry, Cornell University Medical College

ARTHUR L. BENTON

The purpose of this investigation was to make a comparative study of the performances of school children on the Revised Stanford-Binet (4) and on the short *Oral Emergency Test* of Kent (2) It was felt that an examination of the relationships existing between the performances of the same children on both tests might lead to a better understanding of the function of the Kent test in psychometric practice.

Both tests were given by the writer to 55 children who were referred for examination for a variety of specific reasons. In 38 cases, the source of referral was the Children's Clinic of the New York Hospital. Of these 38 cases, 20 were referred for information concerning school placement or because of some physical condition and were not behavior problems. The remaining 18 cases were referred to the Psychiatric Clinic because of behavior difficulties. Fourteen cases were children who were being treated or who were seen in consultation at the Out-Patient Department of the Payne Whitney Psychiatric Clinic because of behavior disorders. None of these children was psychotic Three children were referred by outside agencies for advice concerning school placement and were not behavior problems.

The mean chronological age of the total group (55 cases) was 9 years, 11 months, and the median chronological age was 9 years, 10 months The range of chronological ages was from 6 years, 4 months to 14 years, 3 months The mean chronological age of the "psychiatric" subgroup of children (32 cases) was 9 years, 11 months and the range of chronological ages was from 6 years, 4 months to 14 years, 3 months The mean chronological age of the "non-psychiatric" subgroup of children (23 cases) was 10 years, 0 months and the range of chronological ages was from 6 years, 5 months to

---

*Received in the Editorial Office on October 25, 1937

13 years, 8 months.[1] Twenty-six children were given Form L of the Binet and 29 children were given Form M. The Kent scores were interpolated by a method very similar to that used by Elwood, Burchard and Teagarden (1).[2] In giving the Kent test two scales were usually given. Following the directions of Kent, the scale adopted as giving the subject's score was that scale which brought the subject's rating nearest the middle of the scale.

The test results are summarized in Table 1.

TABLE 1

BINET AND KENT MENTAL AGES AND INTELLIGENCE QUOTIENTS

Test	Mean *MA*	Range of *MA*'s	Mean *IQ*	Range of *IQ*'s
		Total group		
BINET	9 yrs., 3 mos.	5 yrs., 2 mos. to 13 yrs., 10 mos.	94.3	55 to 126
KENT	9 yrs., 1 mo.	5 yrs., 6 mos. to 12 yrs., 9 mos.	93.0	56 to 133
	Non-Psychiatric subgroup		*Psychiatric subgroup*	
	Mean *IQ*	Range of *IQ*'s		
BINET	91.3	55 to 126	96.4	63 to 118
KENT	91.3	56 to 122	94.3	59 to 133

It will be seen that the mean scores on the two tests are almost identical. This holds for the total group and for both the psychiatric and non-psychiatric subgroups. It is interesting to compare these results, obtained from a group of approximately average intelligence, with those of McIntyre and Hoffeditz (3) who gave the Kent test and the original Stanford-Binet to a group of feeble-minded subjects and found that the mean mental age score of the Kent test was 0.8 years *higher* than the mean Binet mental age score To investigate this point further, the scores of the 10 cases in the present group whose Binet *IQ* was less than 75 were examined. The mean Binet *IQ* of this group of 10 subjects was 65.6 while the mean Kent *IQ* was 71 8 The mean Binet mental age of this group was 7 years, 2 months while the mean Kent mental age was 7 years, 10 months The number of cases is small, but so far as the data go, they definitely confirm the finding of McIntyre and Hoffeditz that subjects of low intelligence tend to do somewhat better on the Kent test than on the Binet To check the influence of the scores of the subjects of low intelligence on the total group scores, the mean scores of the 45 children with Binet *IQ*'s of 77 or above were computed The mean Binet *IQ* of this group of 45 subjects was 100.7 while the mean Kent *IQ* was 97 6. The mean Binet mental age of this group was 9 years, 9 months while the mean Kent mental age was 9 years, 4 months. In this particular group, in which the distribution of scores approximates a theoretically normal one with respect to mean *IQ* score (100.7) and range of *IQ* scores (77 to 126), the Binet and Kent scores run quite closely together, but the Kent scores are slightly lower than the Binet scores (3 1 *IQ* points) In order to determine whether this variation was a chance one or whether it represented an actual trend in the relationship of the scores made on both tests, the mean scores of children of higher Binet *IQ* were computed For the 22 cases of Binet *IQ* of 101 or more, the mean Binet *IQ* was 112.0 and the mean Kent *IQ* was 106 8, the Kent score being lower than the Binet score by 5.2 points For the 14 cases of Binet *IQ* of 110 or over, the mean Binet *IQ* was 116 5 and the mean Kent *IQ* was 109.8, the Kent score being lower than the Binet score by 6.7 points For the 9 cases of Binet *IQ* of 114 or over, the mean Binet *IQ* was 119 2 and the mean Kent *IQ* was 111.0, the Kent score being lower than the Binet score by 8.2 points. We find then that with highly intelligent children there is an apparent reversal of the relationship between the two tests which is found with children of low intelligence. In the case of children of low intelligence, the Kent scores run somewhat higher than the Binet score. In the case of a group, the average intelligence of which is normal, the mean scores on both tests are almost identical. In the case of children of high average or superior intelligence, the Kent scores generally run somewhat lower than the Binet scores.

---

[1] The group of children who were examined because of behavior difficulties will be referred to as the "psychiatric" subgroup. Those who were examined for other reasons and who were not behavior problems will be referred to as the "non-psychiatric" subgroup.

[2] These authors write, "In the Kent monograph there is no interpolation made within years. Any score from 13 through 17, for example, on the Low Scale gives a mental age equivalent of seven years; while any score from 18 through 22 gives a mental age equivalent of eight years. Similarly on the other two scales a subject rates a mental age of 9, or 10, or 11, or whatever the case may be but there is no provision made for discriminating *within* any one of these years. In order to make possible comparison with Binet mental ages it seemed necessary to interpolate Kent scores . . . by the method adopted in the present study it was possible to get Kent mental ages of seven years, one month; seven years, four months, eight years, one month, eight years, four months, etc."

In the interpolation of Kent scores devised by the writer, the mental age scores run, in the majority of cases, one month lower than they would if the interpolation method of Elwood, Burchard and Teagarden were used. In the remaining cases, the mental age scores are identical for both methods.

The coefficient of correlation (Product Moment) between the Kent and Binet *IQ*'s was +.84 ($PE = \pm .027$) which is about the same as those found by Elwood, Burchard and Teagarden (+ 88, +.90, +.86) and that found between Binet and Kent mental ages by McIntyre and Hoffeditz (+.87)

Divergence in Scores. Since the correlation coefficient gives only an incomplete picture of the relationships of the scores made on the two tests, the average difference in intelligence quotient rating (disregarding sign) was computed The average difference for the total group was 8.3 points. The average difference for the non-psychiatric subgroup was 7.6 points and that for the psychiatric subgroup was 8 8 points. The difference (1.2) is quite small and is not statistically significant ($\frac{Diff.}{\sigma_{diff.}} = .86$). It may be concluded that, other factors being equal, no greater difference in score on both tests may be expected from children showing psychiatric symptomatology than from those not showing such symptomatology.

The distributions of the differences in score for the total group, the non-psychiatric subgroup and the psychiatric subgroup of children are shown in Table 2.

TABLE 2

PERCENTAGES OF CHILDREN SHOWING VARIOUS DEGREES OF DIFFERENCE IN *IQ* RATING ON THE BINET AND KENT TESTS

	Size of difference (in *IQ* points)					
Group	0-5	6-9	10-13	14-17	18-21	22-25
Total	40%	27%	13%	9%	7%	4%
Non-psychiatric	44%	31%	4%	13%	4%	4%
Psychiatric	38%	25%	19%	6%	9%	3%

As the table indicates, there appears to be little difference between the psychiatric and non-psychiatric subgroups with respect to the distribution of the differences in score. In approximately 67 per cent of the cases the difference in score is 9 *IQ* points or less. In approximately 80 per cent of the cases the difference in score is 13 *IQ* points or less.

In order to determine whether any relationship existed between the intelligence level of the child tested and the magnitude of difference in the scores made on the two tests, the coefficient of correlation between Binet *IQ* and degree of difference between scores was computed. The coefficient of correlation (Product Moment) was +.11, *PE* = ± .09, indicating the absence of linear relationship between intelligence level and degree of divergence in score on the two tests. No relationship between the chronological age of the child and the degree of divergence in score on the two tests was found ($r = 0$). This is in agreement with the finding of McIntyre and Hoffeditz on feeble-minded subjects

SUMMARY

A study of the relationships between the performances of a group of 55 school children on the Revised Stanford-Binet and the Kent E-G-Y Test yielded the following results:

1. For a small group, the average intelligence of which is in the normal range, the mean Binet and Kent scores were almost identical.

2. Children of low intelligence (*IQ* under 75) tended to make higher scores on the Kent test than on the Binet.

3 Children of superior intelligence tended to make lower scores on the Kent test than on the Binet.

4. The coefficient of correlation (Product Moment) between the Kent and Binet *IQ* ratings was + 84, *PE* = ±.027.

5. The average difference in score on the two tests was 8.3 *IQ* points In approximately 67 per cent of the cases, the difference in score was 9 *IQ* points or less.

6. No greater average difference in score on the two tests was found in the case of children showing psychiatric symptomatology than in the case of children not showing such symptomatology.

7. There was no relationship between the intelligence level of the child and the degree of difference in score.

8. There was no relationship between the chronological age of the child and the degree of difference in score.

REFERENCES

1 ELWOOD, M. I., BURCHARD, E M L., & TEAGARDEN, F M An evaluation of the Kent Oral Emergency Test. *J Appl Psychol*, 1937, **21**, 75-84.

2. KENT, G H. Oral test for emergency use in clinics *Ment Meas. Monog*, 1932, No 9. Pp 50.

3. McINTYRE, J. T., & HOFFEDITZ, E. L. Comparative study of the Kent emergency test with feeble-minded subjects. *Train. Sch. Bull.*, 1936, **33**, 22-26.

4. TERMAN, L. M., & MERRILL, M. A. Measuring Intelligence. Boston: Houghton Mifflin, 1937. Pp. 460.

Cornell University Medical College
1300 York Avenue
New York City

Payne Whitney Psychiatric Clinic
New York Hospital
525 East 68th Street
New York City

The Journal of Genetic Psychology, 1938, 52, 401-406

THE RELATION BETWEEN THE INTELLIGENCE OF MOTHERS AND OF THEIR CHILDREN LIVING IN FOSTER HOMES*[1]

Oswego State Normal School

DONALD SNYGG

The primary purpose of this study was to determine the degree to which the intelligence quotients of the parents should be considered in placing a child for adoption, and to explore the possibility of using parents' scores as weights to increase the predictive value of intelligence tests of infants. Many of the more careful child placing agencies have for several years followed policies which prevent or postpone the adoption of children whose parents' intelligence quotients are known to fall below an arbitrary standard These precautionary measures are considered necessary because of the possibility that a large part of the observed similarity (1) in intelligence between parents and their children may be of genetic origin. The findings of Burks (1) and of Leahy (3) that the correlation between the *IQ*'s of children and the mental age of the parents or the cultural level of the home is less in the case of adopted children strongly suggest that environmental factors alone may not account for the resemblance. Leahy (3) found that the correlation between the child's *IQ* and the cultural status of the home was +.18 for adopted children and +.51 for children living in their own homes. Burks (1) found the corresponding correlations between child and mid-parent *IQ*'s to be +.20±.05 and +.52±.05.

Since it is impossible to determine how much of this difference may have been due to heredity and how much to a different parental attitude toward adopted children, the greater age of adoptive parents, or other possible causes, such studies are most valuable for setting an upper limit to the amount of parent-child resemblance in *IQ* which can be assigned to hereditary factors. Within this limit the degree of relationship between the intelligence of children and of parents from whom they have been separated in infancy remains a matter for conjecture.

---

*Received in the Editorial Office on November 4, 1937

[1] From the Mental Hygiene Clinic, Toronto Infants' Home.

The results of the two studies which have made direct attacks upon this problem indicate that the relationship is not large. Lawrence (2) found that in a group of 326 British orphans who had been separated from their parents before the age of one year the correlation ratio between the mothers' social status and their children's Stanford-Binet scores was +.18 for the boys and +.11 for the girls. In the only study so far published (4) on the direct relation between the *IQ*'s of mothers and of their children who had been placed in foster homes, Skeels found that the *IQ*'s of 39 children so placed by Iowa agencies had no relation to the *IQ*'s of their mothers. Data upon the problem are difficult to secure because most agencies have begun the testing of mothers only recently and have used the parents' scores, when known, in the selective placement of their children. Any careful investigation must restrict itself to cases where the placement has been made without reference to the intellectual, educational, social, or economic status of the true parents. Since these facts, when known, are seldom disregarded in placing for adoption, the most fruitful source of information should be from the records of children in orphanages, as in Lawrence's study, or from children in paid foster homes, as in the present one, which is based upon the records of 312 wards of the Toronto Infants' Home.

The Infants' Home is an agency which places and supervises the care in paid foster homes of infants and children less than four years old. The foster homes, by careful selection and close supervision, are quite homogeneous in equipment, social status, and training procedures. The foster parents are, almost without exception, Protestants of British descent, living in residential districts of the city. Most of the foster fathers are salesmen, mechanics, factory foremen, or bookkeepers. In 72 per cent of the homes approved in 1936 the income was between $24.00 and $30.00 a week.

The 312 children who are the subjects of this study differ from the 1350 others who have been examined by the agency's mental hygiene clinic only in that the Stanford-Binet intelligence scores of their mothers are known. Since the clinic makes regular tests of all foster children over one year old, attendance at the clinic does not imply suspected defect. Seventy-nine per cent of the children in this study were placed in foster homes before they were a year old, 90 per cent before they were two, and all before they were four

The scores of the older children were secured through the co-operation of the psychological clinic of the Toronto Children's Aid Society, to which society they had been transferred when placed for adoption or after the age of three. At the time of the last test many of these older children were living in adoptive homes in the selection of which their mothers' scores may have played some part. The majority were living in foster homes quite similar to those maintained by the Infants' Home. Children under three were scored by the Kuhlmann test, those older by the Stanford-Binet. The average *IQ* was 95.17±.54, which closely approximates the average for all children tested by the clinic in 1936. Upon the 16-year basis the Stanford-Binet *IQ*'s of their mothers averaged 78.30±.61. This is probably below the average of all the mothers coming to the agency, as up to 1935 girls who had passed the high school entrance examinations were seldom asked to take the psychological tests. Using the most recent *IQ* for each of the 312 children, the *r* between the mothers' and children's *IQ*'s was +.130±.038. As Figure 1 indicates, there is little possibility of predicting the *IQ* of any child from the *IQ* of its mother, although mothers with *IQ*'s under 70 did have the largest percentage of dull and subnormal children (Figure 2).

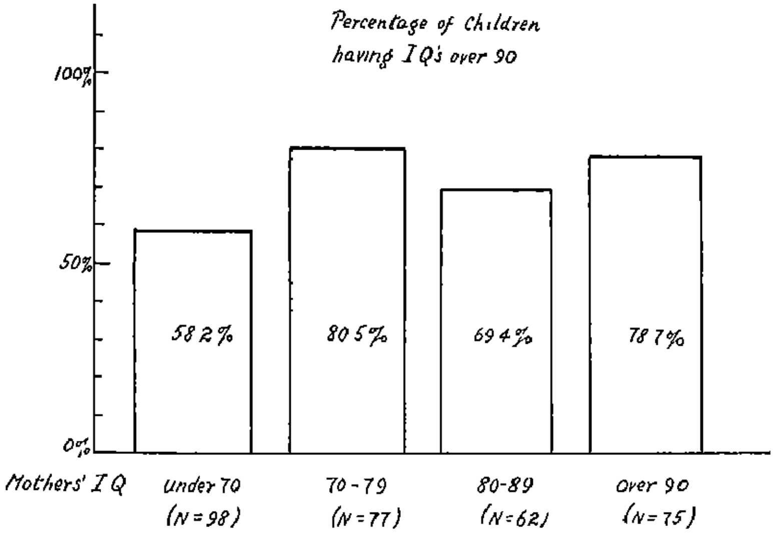


FIGURE 1

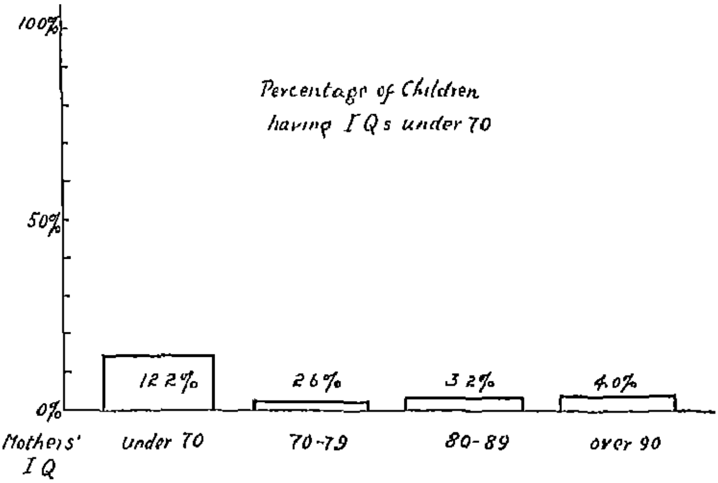


FIGURE 2

Even among this group, however, the majority of the children had *IQ*'s of 90 or higher.

A cursory examination of the results indicates that there is no straight line correlation, and the correlation found is largely due to the low scores made by the group of children whose mothers have *IQ*'s under 70 (*MA* less than 11-2). The average of these children was 90.74±.54, compared with 97.43±.52 for the children of mothers with higher *IQ*'s. Rather surprisingly, the best showing on every count was made by the children of the 77 mothers whose *IQ*'s were between 70 and 79 These mothers not only produced a lower percentage of subnormal (under 70) children than any other group but produced 17 children with *IQ*'s over 109, as many as all the 235 other mothers combined. Even the mothers with *IQ*'s under 70 produced as high a percentage of bright children as did the mothers with *IQ*'s of 90 or higher.

It would appear that among this group of children knowledge of the mother's *IQ* has very little predictive value. If a chance selection were made from the children whose mothers had *IQ*'s over 89, the chances would be 79 out of 100 that the child chosen would have a normal (90 or over) *IQ*. If the selection were made from all children whose mothers had *IQ*'s over 69 the chances would still be 77 out of 100; if the choice were made from among all the children studied the chances would be 71 out of 100. On the other hand, the chances of securing a child with an *IQ* of 110 or better would be poorer (6.6 out of 100) among the children of the higher *IQ* mothers than among the whole group (10 9 out of 100).

It will be remembered that these results are based upon tests upon children of all ages. It is necessary to consider, therefore, the possibility the inclusion of test scores by young children may have masked a true correlation greater than the + 13 we have found If this were true the scores of the older children should correlate more highly with the scores of their mothers Accordingly the correlations for each age group were calculated independently. As Table 1 indicates, there was no increase in correlation as the children became older. The correlation between the *IQ*'s of the 70 children who were over 5 years old when tested and the *IQ*'s of their mothers is +.120±.078 As has already been noted, there is a possibility that selective placement of some of these children may have helped cause this correlation.

TABLE 1

Correlations between the *IQ*'s of Mothers and Children

Age of child	r	N
Under 1 year	+ 115	31
1 to 1½ years	+ 082	118
1½ to 2 years	+ 079	109
2 years	— 119	128
3 years	+ 153	96
4 years	+ 045	52
3 years or more	+ 086	149
4 years or more	+.129	93
5 years or more	+ 120	70
5 years or more*	+ 116	53

*Placed in foster homes before the age of two

Unfortunately, Stanford-Binet scores of the children's fathers are not available. A sampling of the case histories, however, indicates that they are not superior to the mothers in education or economic status. Assuming the correctness of + 13 as the child-parent *r*, and assuming that the scores of the parents are uncorrelated, the most probable child-midparent *r* would be +.19, much too low for any use in prognosis.

As might be expected from the low correlations involved, a combination of the mother's and infant's scores is of no value in predicting the child's later *IQ*. Correlating the average of the one-year *IQ*'s of 32 children and the *IQ*'s of their mother's with the *IQ*'s of the children on their recent tests (ages 3 to 8) the *r* was +.275. Correlating the children's *IQ*'s alone the *r* was +.405.

SUMMARY

The correlation between the *IQ*'s of 312 children living in foster homes and the *IQ*'s of their true mothers was +.130. The correlation does not increase with age. The *IQ*'s of 70 children over 5 years old correlated +.120 with the *IQ*'s of their mothers. Part of the correlation may be ascribed to selective placement.

CONCLUSIONS

1 There appears to be a slight correlation between the Kuhlman or Stanford-Binet scores of children and the Stanford-Binet scores of their mothers from whom they have been separated in infancy.

2. The correlation is too low to warrant the use of a mother's *IQ* in predicting the future rating of her child reared in another home.

3. Predictions based upon consideration of both the mother's and child's scores are less reliable than those based upon the child's score alone.

REFERENCES

1. BURKS, B. S. The relative influence of nature and nurture upon mental development. *27th Yearbook Nat Soc Stud. Educ*, Pt I, 1928, 219-316.
2. LAWRENCE, E. M. An investigation into the relation between intelligence and inheritance. *Brit J. Psych*, *Monog. Suppl*, 1931, **16**.
3. LEAHY, A M. A study of adopted children as a method of investigating nature-nurture. *J Amer. Statis Asso.*, 1935, **30**, 281-287.
4. SKEELS, H M. Mental development of children in foster homes. *J. Genet Psychol.*, 1936, **49**, 91-106.

Oswego State Normal School
Oswego, New York

The Journal of Genetic Psychology, 1938, **52**, 407-423

AN INVESTIGATION OF READING RETARDATION*

Department of Education, Johns Hopkins University

Arthur Lichtenstein[1]

This investigation was primarily an attempt to determine the specific causes of reading difficulty in a group of retarded readers in a public school system From the results of group intelligence and reading achievement tests, all cases who were three years or more below mental age, and also below the 4th grade in reading achievement, were noted This produced a list of about 40 names.

The schools in which these children were placed were visited and each case investigated further, with a view to detecting physiological defects which might be related to the reading disability Lefthandedness, mirror writing, and speech defects were the principal things sought in this investigation, with further data of interest obtained by interviewing the principal, the teacher and the child. From the group of 40, by this process, 20 cases were selected in which the disability seemed most likely to be related to physiological or psychological causes. Of these, 18 were boys, 2 were girls. This proportion seems consistent with most findings of students of reading difficulties

These cases were given the diagnostic reading tests, which included the Gates *Diagnostic Tests* and the Betts *Ready-to-Read Tests.* The latter are used chiefly to detect physiological difficulties in visual perception such as muscle imbalance or ametropia, while the former tests phonics, auditory memory, associative learning, etc. Each case was then analyzed completely on the basis of the tests, and specific recommendations for remedial work were made.

In addition, a brief statistical analysis was made of the 20 cases, which constituted a fairly homogeneous group Some of the results of this analysis are presented below It must be borne in mind, of course, that 20 cases are too few upon which to base any conclusions,

---

*Accepted for publication by Buford Johnson of the Editorial Board, and received in the Editorial Office on November 5, 1937

[1] The author wishes to express his indebtedness to Dr. Harry F Latshaw, Director of Special Education in Baltimore, at whose instigation the project was undertaken.

but this work may be considered the foundation for a more comprehensive study to be done in the future.

Referring to Table 1, we find that the group of 20 physically normal retarded readers are slightly subnormal in intellect, on the basis of both individual and group tests. Rather oddly results show that this group, handicapped in reading, does better on the group tests, which generally involves a much greater amount of reading. However, the group tests used in this study were in every case the most recent available, since there was no one test on which every case or almost every case had made a score. Thus, some of the tests are Pintner-Cunningham, which involves no reading, and some are Illinois which involves a very considerable amount of reading. In general, it is safe to assume that the group tests involve more reading than the Binet and the difference in the median *IQ*'s is surprising, to say the least.

TABLE 1

MEDIANS AND MEAN DEVIATIONS OF SCORES IN CERTAIN SECTIONS OF THE DIAGNOSTIC READING TESTS

Subject	Median	M.D.
Binet *IQ*	83	10.5
Group Test *IQ**	91	7.5
Chronological Age	147 mos	17.5
Binet Mental Age	119 mos	17.5
Grade Placement**	4*B*	1.1 grades
Handedness Ratio***	1.33	0.18
Gray Oral Reading****	2.3	0.9
Gates Group Tests (average)	2.8	0.7
Phonetic Abilities (average)	2.1	0.7
Auditory Perception	2.3	1.1
Visual Perception	4.3	1.7
Associative Learning	5.0	1.3
Auditory Memory Span	3.2	1.0

*Most recent for each case.

**This is very approximate, due to the fact that many were in special classes. The achievement test scores were used as far as possible.

***$\frac{L}{R}$ scores on the three tests of handedness; a very crude index. See below.

****From this point to the end of the table, scores are in terms of grade levels.

The median chronological age of these pupils was 12 years, 3 months, with a large mean deviation (nearly one-and-one-half years). The median grade placement was 4*B*. As stated in the table, this is a very crude approximation, since it was not feasible to assign accurate grade levels to children in Special Centers, Opportunity Classes, or Pre-vocational Schools. The median grade placement for each of these groupings, based on achievement test results, was used in making the classification. Assuming 4*B* to be about right, the average retarded reader seems to be three years below normal for chronological age on grade placement

The median mental age of these cases is considerably below the *CA*, being 9 years, 11 months On the basis of *MA* a placement of 4*B* would be about a half year below what would be expected This might be accounted for by the specialized reading disability.

The median handedness ratio is 1.33, but this is quite meaningless, since the cases were selected to some extent on the basis of handedness (see second paragraph). Of the 20 cases four wrote left-handed and three threw left-handed One case wrote left and threw right but the ratio (0.74) indicated the left hand to be dominant. The investigation of handedness was not very comprehensive. A dynamometer was used and relative strength of grip of the two hands determined Then the *X* and *O* tests prescribed in the Betts *Manual* were given to measure finer coordination. A combined ratio was given from these three tests.

Tests of eye dominance were also modelled after Betts except that all instruments were not available, and usually only one or two were used—the manoptoscope and the light. The indices obtained were not felt to be highly reliable. On many occasions also one eye is wholly dominant resulting in a ratio like $\frac{6}{0}$ or $\frac{0}{6}$ Results of this type do not lend themselves to mathematical treatment. Therefore the results of the eye-dominance tests are not included in Table 1, but the results for each individual were considered in making recommendations for his case.

The medians and variability for some of the diagnostic tests are also presented in Table 1. The median oral reading level for the group was about middle of 2*B*, or two full grades below placement Individual types of errors in this test as in the others given in the table were fully discussed in the case reports In silent reading these children seemed to do slightly better—by one semester in fact, as measured by average grade levels on the Gates *Silent Reading Tests*. This in understandable—memorization and guessing ability enter into the *Silent Reading* scores more than in the oral, and phonetic defects, of which there were many in the group, would not be such a handicap.

In phonetic abilities the median grade level is lowest of any in the list. This seems to be significant and indicates that these children, or children with these types of defects, are likely to profit by more intensive phonetic training than they have, for the most part, been receiving. It must be noted that the results of the phonetic abilities tests are based on two different types of test, due to the fact that the older edition of the Gates test was used during the first part of the study and the new edition during the latter part when the supply of copies of the original was exhausted The two tests differ considerably in their set-up, the latter being far more elaborate. Differences between the two editions caused other difficulties in analyzing the results of the visual perception tests, the associative learning, and others, which were modified in some degree in the new edition

It is interesting to observe that the median visual perception level of this group is two grades above the median auditory perception. This difference appears to be large enough to generalize the rule that the visual approach would be the most efficient technique in coaching these retarded readers. There are, of course, exceptions, which are noted in the individual case reports

The associative learning ability of these youngsters is surprisingly good, as shown on the tests they were given. They achieved a median grade level of 5.0—a grade above their median placement, and up to their median mental age. Some individuals proved to be superior in visual-auditory learning, others in visual-visual learning, and these findings are presented in the individual reports, and recommendations for remedial work are based on them.

The auditory memory span is inferior to the learning ability, which seems to indicate that the learning is not memorization, as has been suspected, and is in fact pretty well known to be the case for many non-readers.

Table 2 presents some of the correlations determined from the analysis of the test results. The small number of cases reduces the significance of these results; and in addition some of the factors compared were not exactly the same, due to use of the revised form of the Gates test in some instances as explained above. The cor-

TABLE 2

SOME CORRELATIONS DETERMINED FROM THE RESULTS OF THE DIAGNOSTIC READING TESTS*

	Relationship between.	Coefficient	PE
1.	Binet and Group Test *IQ*	—.07	08
2	Chronological Age and Grade Placement	31	14
3	Mental Age and Grade Placement	51	.12
4	Mental Age and Gray Oral Reading	61	11
5	Mental Age and Associative Learning	.11	17
6	Mental Age and Phonetic Abilities	40	.14
7	Mental Age and Gates Reading Group	.29	.17
8	Mental Age and Visual Perception	.06	18
9	Mental Age and Auditory Memory Span	25	17
10	Visual Perception and Auditory Memory Span	— 08	18

relation coefficients presented, therefore, may be regarded as very crude indices, pointing to some interesting possibilities for further investigation, but of little importance in themselves. Probably the most startling is the negative relationship between the Binet and the group mental tests. The very low relationship between mental age and associative learning ability which are in many quarters considered synonymous, is somewhat unexpected. The highest coefficient in the table seems to indicate a positive connection between mental ability and oral reading ability as measured by the Gray *Test*. This is by no means surprising, but it is somewhat unexpected to find this coefficient much higher than that between mental age and the group reading tests for example. However, a glance at the probable errors of all these figures reveals the extreme unlikelihood of more than one or two being at all significant

The chief contribution of such investigations as this is in the individual adjustments that are made; extracts from several sample case reports are appended to give some idea of the nature of the analyses.

KENNETH O

1. *Kenneth O* was referred on the basis of reading achievement consistently far below mental level. He was first tested at the age of 11-10, in grade 4*A*. Previous group test results were·

Age	Test	Score	(*IQ*)	Grade level
6-0	Detroit Intelligence	19	(96)	
6-7	Pintner, Cunningham Intelligence		(89)	
8-0	Detroit Word Recognition	14		
8-7	Gates Primary Reading	(average)		2 6
9-7	Gates Silent Reading Type A			2A
10-0	Stanford Achievement	38		3A
11-7	National Intelligence	52	(76)	

He was given the Stanford-Binet Test, *CA* 11-9, *MA* 10-4, *IQ* 88. Basal age 9 years, scatter 6 years. Passed clinical third-year reading test The results of the diagnostic reading tests are given in Table 3

TABLE 3

DIAGNOSTIC READING TEST RESULTS, KENNETH O

Gates Diagnostic Tests	Score	Grade	Age
Pronunciation Test	51.5	2 4	7.9
Phonetic Test			
Reading capitals	13	4 4	10 2
Errors.			
Reading small letters	14	4 2	10 0
Errors'			
Set A			
Sounding letters	17	2 2	7 7
Sounding syllables	30	1.8	7.3
Sounding compound syllables	5	2.9	8.4
Set B			
Giving letters for sound	7	1 8	7 3
Spelling syllables	3	7 1-1 6	Below
Spelling non-sense words	2	1.8	7.3
Spelling words	11	3 0	8.6
Visual Perception			
Perception, Figures	15	3.0	8.6
Perception, Digits	32	3.0	8.6
Perception, Words	35	3 0	8 6
Selection, Figures	16	4.6	10 5
Selection, Words	17	3 0	8 6
Visual Analysis and Recognition			
Geometrical designs	10	Far above norms	
Word-like character	11	Far above norms	
Visual Memory Span			
Span of figures	3	7.1-1.6	Below
Auditory Function			
Auditory discrimination	6	4 25+	10 04
Auditory Memory Span			
Digits	6	4.5	10 4
Letters	6	11 4	5.5
Non-sense words	5	5 75	11.65
Words	5	4 5	10.4
Associative Learning			
A^1 Visual-visual	9	Above norms	
A^2 Visual-visual	8	5.25	11 05
A^3 Visual-visual	5	2 25	7 75
B^1 Visual-auditory	7	3 6	9.3
B^2 Visual-auditory	10	Above norms	
B^3 Visual-auditory	9	Above norms	
Gates Primary I			
II	34 7	3 55	9 25
III			

Betts Ready-to-Read Tests

1. Visual Readiness. Normal, slightly inferior in word forms
2. Auditory Readiness Poor in fusion and span, otherwise normal.
3. Visual Sensation and Perception: Normal.
4. Oculomotor and Perception Habits: Normal at about second grade level. Many letter reversals noted, particularly at third grade level.

Tests of Handedness and Ocular Dominance.

Right hand dominates, ratio 1.46. He throws and writes with the right. The eyes register mixed dominance.

Observations by Teacher.

No bad habits in silent reading. In oral reading, he hesitates, phrases improperly, mispronounces, guesses, substitutes, reverses letters and reverses words. There is no evidence of eye-train, but the attention wanders considerably. There is no speech defect. Motor control is satisfactory except for the handwriting, which is considered poor. There are no emotional difficulties. His best work is in Arithmetic, poorest in Reading and Spelling He is retarded two and a half years. Attendance good, home attitude unknown, probably not over-cooperative, as Principal reported that mother wanted to know why Kenneth had to be tested when none of the other children were.

Conclusions and Recommendations.

There is apparently no physiological basis for his reading retardation. Nor is his capacity inferior based on test results as well as general impression. His phonetic ability is low, especially in sounding phonograms. His visual perception is around third grade level. These two are the weakest points and remedial work should center around them. Flash cards and work with families of sound-syllables should be helpful. Then the recognition of words in context can be worked with. He uses the spelling method in attempting new words and this gives him no help His chief weakness seems to be with consonantal sounds. His attempts show a marked lack of self-confidence which probably needs bolstering. This should not be difficult as he seems to be a stable, well-balanced youngster, rather suggestible, and very responsive to motivation.

Joseph S.

Joseph S. was referred because his reading test scores were consistently below his mental level. He was first examined at the age of 12 years, 2 months, in the 3*A* grade.

On the Stanford-Binet test given nearly three years before, with a chronological age of 9 years, 11 months, Joseph measured 8 years, 2 months mentally, thus having an *IQ* of 82. He based at 6 years and scattered through the ten-year test He failed the Wallin *Reading Test* at the first grade level, but turned in an exceptionally good performance on the form-board tests. The physical examination by the school was negative. Now he is reported by his teacher to be slightly deaf. He has a scar from tracheotomy. Vision in both eyes is normal.

Joseph was reported as talkative, restless, indifferent, and unambitious At the time he was first examined he was in grade 2*A*, but doing 2*B* reading and arithmetic and 1*A* spelling. His conduct was rated very poor

The investigation of home conditions and family history revealed that Joseph was of German-American parentage and his birth and development were normal. At the age of four, he had an extremely serious attack of black diphtheria, at which time his wind-pipe was cut and a tube was inserted for feeding. He was apparently much spoiled since that time and developed into a considerable problem as a result. The mother seemed to be a poor disciplinarian.

Joseph is the second of three brothers He plays with boys in the neighborhood and his younger brother, and seems to get along quite well with them. In general, he was said to be a jealous type, crying readily and easily hurt or upset.

A Pintner-Cunningham *Intelligence Test*, at the age of 8-0 gave Joseph an *IQ* of 97. The Gates *Primary Reading Tests* at the age of 10-4 resulted in an average reading level of 2*B*. He scored 19 on the *Metropolitan Achievement Primary Reading Test*, equivalent to grade level 2.8 at age of 11-11. At this time he was in grade 3*A*. A semester later he raised his level on this same test to 4 1.

Results of Betts Ready-to-Read Tests

Joseph was given the Betts tests first. Tests of visual readiness revealed as notable characteristics of his reading extreme slowness and some letter reversals, such as *d* for *b* and *p* for *q*. His auditory readiness revealed a marked weakness fusion and imperfect perception. His auditory acuity was somewhat below normal, and he was referred to an otologist.

Tests of visual sensation and perception were negative. The visual acuity is slightly inferior but not enough to affect his reading. In all other respects his eyes functioned quite well.

In the realm of oculomotor and perception habits he began meeting difficulties. The words at second grade level offered obstacles: he read pen for pin, house for horse, etc., scoring 16 out of 20 in gross binocular vision at that level. At third grade level he scored 12. There were no marked differences between the right and the left eye. His performance with superimposition was slightly inferior and he had some difficulty with the stereopsis test. There was no consistent indication, however, of muscular imbalance as the cause of his difficulty. It seemed more likely to be in the field of phonetic analysis. He spelled out words without getting the sound at all.

He was tested for eye and hand-dominance and the right was found dominant in both cases. His teacher stated that among his reading habits were some lip movements, unnecessary head movements, hesitation, mispronunciation, guessing, omission, and marked letter-reversals. This latter difficulty was noted by the examiner; it seemed quite pronounced only in two places. The teacher had noted also poor attention habits and a speech defect (this was also noticed by the examiner—it appears only at certain times and is a sort of wheeze. There is also a slight lisp.) She reported his best work as being in Arithmetic, his poorest in Reading, attendance good and home attitude very cooperative.

Results of Gates Diagnostic Tests.

Four months later, Joseph was given the Gates *Diagnostic Battery* which appeared to point out quite clearly the nature of his specific difficulties.

The *Gray Oral Reading Test,* given first, resulted in a raw score of 14, corresponding to a grade score of 3.7. He made several insertions and omissions of parts of words, and seemed to get his cues from initial letters fairly well but to miss in the middle and at the end of words. As the paragraphs grew more difficult his refusals increased rapidly. Interestingly enough, the Gates *Oral Context* test produced exactly the same grade-score as the Gray

Mispronunciations, including reversals of parts, wrong beginnings and wrong endings, were the chief sources of error.

On the *Graded Word Pronunciation Test*, which offers no opportunity for contextal cues, Joseph did not do quite so well—his raw score was 56.5, equivalent to a grade score of 3.4.

Tests of *Perceptual Orientation* followed. On *Isolated Words*, Joseph scored 12—grade score equivalent to 2.8. He made no reversals, which this test is particularly designed to detect. His errors were in both consonants and vowels In *Word Recognition from Visual Presentation* he made a perfect score—no errors. From *Auditory Presentation* he made four errors· two of wrong ending and two of wrong orientation (reversals).

Visual Perception Techniques revealed consistently low ratings. In all types of phonogram combinations he was considerably below third grade level He did most poorly with consonant phonograms, consonant-vowel phonograms, blending and letter sounds He can name the letters correctly, although he is extremely slow with them. However, when it comes to sounding the letters, he falls down miserably, recognizing satisfactorily only *o*, *i*, *e*, and *a*. There is clearly almost no phonetic ability present.

In auditory perception, his techniques are not superior, ranging from grade level 1.3 to 2.25. He appeared slightly more skilful at giving the letter for a sound than the reverse, but only slightly He was better with diphthongs than with either consonants or vowels!

In auditory discrimination, Joseph did quite well—reaching grade level 3.0 in repetition of nonsense words, and making a perfect score in distinguishing words If there is any defect in auditory acuity, it certainly did not hamper him in these tests.

In the tests of visual perception, which check ability to distinguish small differences in figures and numbers, he did excellently, ranking about sixth grade level.

In spelling words he was at 2.55, in writing spelled words 2 8. In writing letters from dictation he also did very poorly. There are no norms for this test, but it revealed again an insufficient knowledge of his letters.

Tests of associative learning produced interesting results On the visual-visual test, with easy figures, he made a perfect score, equivalent to seventh grade level or better. On the visual-visual test with word-like figures he fell down badly, making a score of 4, grade level 1 75. On both tests of visual-auditory learning he made perfect scores The chief difficulty in this realm seems to be in making the transfer from simple to complex symbols, such as those involved in a word.

On the basis of this showing, Joseph should be able to learn the conformation of the letters, and then drill in the "gestalts" of words should be effective The visual modality seems likely to produce more success than the auditory in forming associations between words and letters, if the tests of perception are significant.

Tests of memory span concluded the session. For digits, Joseph's span is 5, for letters, nonsense syllables, and familiar words, it is 4 The average grade level for memory span is 3 25, slightly below his grade placement and far below his age level

Conclusions and Recommendations

Grouping the tests into specific fields of reading readiness and ability, we find that in methods of working out recognition, pronunciation and meanings of words, his average grade level is 3 6. In perceptual orientation, he is above third grade On both these tests the purport of the results seems to be that recognition of phonograms and letter sounds is extremely weak, but the reading habits such as eye-movements, etc , are established. The visual perception techniques average 2 1 grade level, bearing out the conclusion that visual recognition of symbols and association of symbols with sounds and meanings, are defective in this boy The Betts test indicates no physiological factor responsible for this, the connections simply have not been made. Auditory perception techniques are no better established, averaging 1 8 Yet auditory discrimination seems perfectly normal, as does visual perception. The one outstanding defect in associative learning has already been pointed out, and would seem an excellent focal point to begin remedial work, viz , making the transition from associations of simple characters or symbols with meanings and sounds, to more and more complex associations of the same type The technique of auditory fusion is particularly defective and needs considerable attention.

Individual attention in methods of perceiving and analyzing words should be helpful in Joseph's case. Arranging words to show similarities and differences, calling attention to unique and common elements, the kinaesthetic method of saying the word and writing it at the same time; use of phrases and sentences in drill to give contextual assistance, flash cards, training in syllabication, are all techniques of coaching which have been found very helpful in such cases as this. Once the fundamental factor in phonetic analysis, i.e., what distinguishes one sound or one symbol from another, has been grasped, Joseph should be able to make fairly rapid progress. His intellectual limitation is probably not so great as the *IQ* indicated, since the reading defect probably enters into it. It will be interesting to give another Binet test after the remedial work has progressed for some time, to see whether there is any gain. If there is none, it must be realized that Joseph is not apt to make very rapid progress even with individual attention

CHARLES J

Charles J. was referred with a lengthy history of reading and speech difficulty He was 14 years, 4 months old, and in the 6*B*. He had moved a good deal in his early school career, attending schools in three states.

His home situation was favorable, with intelligent parents, who were reasonably strict with the boy. The grandmother, who lived with the family, was reported to humor the boy at times. The neighborhood was rated good and the home well-kept. Charles' playmates appeared to be younger than himself and he is said to have screamed when attacked by older boys, not to be able to stand rough play. He apparently got along satisfactorily with younger boys. It is also noted that the shock of an ordinary fall affected him so that he had to go to bed for the whole day. He was afraid of dogs. He talks in his sleep and occasionally screamed at night when an infant. He is generally good-natured but has a temper.

His developmental history reveals normal birth, weight nine pounds He was breast-fed for two months, bottle-fed for seventeen months. He was in poor health as a baby. He had spasms at nine months (during a trip), whooping-cough and measles at three years, and *serious pneumonia at four years* From the latter illness he spent several months recovering. He talked and walked late—took his first unsupported steps at 19 months, and first used sentences at four and a half years.

The general picture is that of a nervous, oversensitive, timid child, with a neurotic constitution complicated by marked physical defects. If not for the unfavorable developmental history, it might be suspected that the pneumonia at age four was actually encephalitis, which is often so wrongly diagnosed. His behavior shows some signs of post-encephalitic disorders, but apparently he had the peculiarities before the pneumonia. It seems probable also that the psychiatric diagnosis would have included this possibility had their examination revealed any likelihood of its presence.

In addition, it should be noted that Charles is the youngest of three children His older brother and sister are doing *excellent* work in school.

A Stanford-Binet *Intelligence Test* at the age of 8 years, 3 months gave him an *IQ* of 107 He based at seven and scattered through ten. Comments on the test were that he was left-handed but was taught to write with his right hand. His enunciation was clear, but speech labored and muscular coordination poor Physical defects found were: overweight, stained and carious teeth, astigmatism, eye-strain and blepharitis He was wearing glasses at the time His educational record noted him as cheerful, talkative, friendly and popular, but his attention was poor. His teacher stated that although he was extremely slow and rarely completed any work he undertook, Charles seemed quite satisfied with his work

At the age of 9-0 he was seen by a psychiatrist, who reported that he based at four on the *Intelligence Test* and scattered through nine, making an *MA* of barely eight years, and an *IQ* of 88. The physician suspected that part of the boy's progress to date was due to help from his mother. Difficulty of determining handedness due to very inferior coordination was also pointed out in this report.

A month or so later, he was given the *New Stanford Reading Test,* Form V, and scored 43, corresponding to a grade score of 3 6. Since he was in *3A* at the time, he might be said to have been at grade level One month later he was given the *New Stanford Reading and Arithmetic Test,* Form X, and revealed a marked gain He was being tutored privately at the time His reading grade score was 4.25, arithmetic 4 5 His spelling was also tested and found to be at grade level 3.7.

Shortly afterward Charles was transferred to a private school, where he remained a year and a half. He returned to public school in grade seven, and was re-adjusted to *6B,* where he failed of promotion the following February. He was then given the Stanford-Binet again. At chronological age 12-3, he scored a mental age of 14-1, *IQ* 115 He based at 9, failed on Designs and 60-Words at year 10, and scattered through 18, passing Digits Forward and Backward at the superior adult level. Some of his failures, particularly those in year 10, may be attributable to his poor co-ordination and slow speech. His auditory memory is outstanding. In Form-boards his performance was very poor.

His educational record at this time included the report of an *Otis Intelligence Test,* on which he attained an *IQ* of 96 His *IQ* on the *Pintner-Cunningham Test* given in the first grade was 89. His lower rating on group tests points to a reading difficulty as the cause, since reading is decidedly a factor in such tests Achievement tests had also been given His grade level on the *Metropolitan Reading Test* was 6 1, and on the *Arithmetic* was 5.8. He was clearly not achieving up to his capacity. He was ranked toward the lower end of the scale in application, interest, attention, memory and comprehension. On the *Social Adjustment Score Card,* he makes a total of 25, which indicates about average adjustment. In leadership, industry, courtesy and obedience, he is about average. In honesty and truthfulness he is rated high, in neatness and carefulness he is low.

After his failure he was given Form V of the *New Stanford Reading and Arithmetic Test.* His reading level was 5.8, arithmetic 6.7. His arithmetic computation was 1.5 grades higher than his arithmetic reasoning.

Results of Betts Ready-to-Read Tests.

Charles was seen at his school by the examiner He had no difficulty with the tests of *Visual Readiness,* making perfect scores on each. These include Letter Forms, Word Forms, Phonetic Elements and Recognition of Letters. Third grade children are expected to make perfect scores on these tests.

On the *Auditory Readiness Tests* he did about as well. On auditory span he scored 24 of a possible 25, an exceptional performance which checks with his fine showing in auditory memory on the Binet His auditory acuity was not normal, but since he had a bad cold at the time, this need not be considered seriously, as his physical examinations all concur in reporting his hearing as adequate. Auditory fusion, perception of consonant sounds, and perception of vowel sounds, were revealed as perfectly satisfactory. His defects clearly had to be sought in other fields

The tests of visual sensation and perception were given next. He was given the tests with and without his glasses His binocular vision was normal, that is to say, he apparently sees with both eyes. However, his far-point fusion and near-point fusion were unsatisfactory His fusion power is evidently at a low level. His glasses do not correct this defect, which is muscular in its nature. This alone would account in some measure for the reading difficulty.

The visual acuity tests were given next. With his correction Charles' binocular acuity was 90 per cent, left eye acuity 70 per cent, right eye 90 per cent Without the correction, his binocular acuity was 80 per cent, left eye 60 per cent, and right eye 80 per cent. Clearly the glasses effected considerable improvement, bringing the vision from what corresponds to as low as 20/75 on the Snellen *Chart* to practically normal with both eyes The left eye is still inferior, however, even with the correction corresponding to 20/60 on the Snellen Chart. This also is a frequent cause of reading difficulty.

The test of vertical imbalance also revealed a marked defect. A curious form of hyperphoria was shown, the deviation being slightly below with the correction worn, and slightly above without it.

Charles' stereopsis level was 100 per cent both with the glasses and without them.

The test of lateral imbalance indicated another visual defect. With this correction Charles showed a tendency to exophoria (divergence). Without the correction it was toward esophoria (excessive convergence). The reflexes of accommodation and convergence are not normal in either case

The tests for ametropia or refractive errors were given next The vision was abnormal both with and without the correction, but the glasses corrected the abnormality to a considerable extent With the correction, however, there was some evidence of hypermetropia in both eyes Without the correction, the picture was that of myopia There was evidence of astigmatism in both eyes without glasses and some slight evidence of it in the left eye even with the glasses.

Tests of *Oculomotor and Perception Habits* followed. Charles first read the words using only his right eye, both with and without his *correction. He was clearly suppressing the vision of the left* eye when put into a reading situation, whereas he did not do this, although *the left eye was shown to be more defective,* in the non-reading (picture) situation This is an interesting habit. It was borne out in both tests of gross binocular vision.

In the letter orientation test, he did the same thing when not using his glasses, but using them he did read the words opposite his left eye, thus making a much better showing. In the tests of number and word orientation he made practically perfect scores. These are tests of superimposition without depth perception. In the stereopsis test he reported no depth perception without his glasses at first but later stated that he had it With glasses he reported depth perception throughout, making very good scores.

Charles was next given tests for eye dominance. The right eye appeared markedly dominant, the ratio being 5.0 This was expected in view of his showing on the perception tests.

Tests of handedness were next given. These included the Smedley Dynamometer for strength of grip and *"X"* and *"O"* tests for coordination. On the dynamometer, the left hand was *very* slightly superior, but on the other tests the right hand showed up much better. The combined ratio was 1.27, indicative of right hand preference.

Conclusions and Recommendations.

On the basis of these tests, the nature of Charles' reading difficulty *appears wholly physiological, probably tied up with a* neurological mechanism that is markedly defective in most of its aspects. *His present teacher reports that he frequently frowns or grimaces* in class, a habit which she thinks may be due to the effort of trying to talk. She also noted other evidences of poor motor coordination: his handwriting is extremely poor, his art work is bad, he runs awkwardly and stiffly, his speech defect is quite apparent in class. She also stated that he appeared to be entirely lacking in emotional expression. This was not entirely justified in the interview with the examiner, although the general sluggishness of his responses might well give that impression.

It is possible that with intensive work in the correction of the muscular defects his reading can be considerably improved. No doubt better results will be obtained in remedial work if the appearance of a reading situation is avoided, since certain habits appear to depend upon the situation being non-reading. He is being referred to an ophthalmologist for this purpose.

From the temperamental side it would certainly seem advisable that Charles should be "pushed" in school as little as possible. The tendency which is prevalent in many families of continually setting the example of superior brothers and sisters before the black sheep can do much harm in the case of a nervous, awkward boy like Charles. Allowed to proceed at his own pace, with his good intellectual endowment, he may eventually overcome the worst of his handicaps unaided. The help he *can* get should, of course, contribute to this desirable goal

SUMMARY

There is some evidence that individual analysis of children with marked reading defects does a great deal to improve their status. Three important instruments used in this investigation were the Stanford-Binet *Intelligence Tests,* Gates *Reading Diagnosis Tests* and the Betts *Ready-to-Read Tests.* On the basis of findings of these tests individual recommendations were made A checkup of many cases indicated that the difficulties had been correctly diagnosed. A complete check has not yet been made. Excerpts from the individual case reports of three subjects are included.

REFERENCES

1. BETTS, E A Betts Ready-to-Read Tests, Manual of Directions Meadville, Pa Keystone View Co., 1934.
2 ——— Reading disability correlates *Education,* 1933, **66**, 18-24
3. BOND, G L Auditory and speech characteristics of poor readers *Columbia Univ, Teach Coll Contrib Educ,* 1935, No. 657
4 FENDRICK, P Visual characteristics of poor readers *Columbia Univ, Teach. Coll. Contrib Educ,* 1935, No. 656
5. GATES, A I. Improvement of Reading (Revised) New York Macmillan, 1935.
6. ——— Viewpoints underlying the study of reading disabilities *Elem Eng. Rev.,* 1935, 12, 85-90.

Department of Education
Johns Hopkins University
Baltimore, Maryland

SHORT ARTICLES AND NOTES

The Journal of Genetic Psychology, 1938, 52, 425-428.

OBSERVATIONS OF DOMINANCE-SUBORDINATION IN CATS*

Department of Psychology, Brooklyn College

CHARLES NILSON WINSLOW

Seldom are cats observed in large numbers, so that it is not surprising that little has been written concerning their social behavior. In animal laboratories, however, such opportunities are sometimes afforded. It was in the living quarters of a cat colony in the Laboratory of Neuropathology of the Psychiatric Institute in New York City, that the writer made the observations he will discuss in this paper.

Traditionally the cat is well-known for its predilection for solitude, and its relative self-sufficiency. Any contacts with other cats seem peculiarly accidental and transitory. It is indeed the exception rather than the rule for adult cats to run and play with each other as do dogs. When placed together in large numbers within one room, however, opportunities for more intimate social contacts are made possible. The writer concerned himself first with dominance-subordination behavioral patterns of cats to see how they compared with those reported for primates by Zuckerman (8) and Maslow (1, 2, 3), and Murchison (4, 5, 6) for domestic fowl, and second, to see how dominance-subordination patterns are related to other forms of behavior.

Within a few minutes of observation of the behavior of the cats in the colony it became apparent that communistic social relationships did not exist. A swaggering and strutting male was seen roaming around the room, from whom most of the other animals withdrew. He was not the largest animal in the cage, but possessed

---

*Accepted for publication by C. J. Warden of the Editorial Board, and received in the Editorial Office on July 22, 1937

the "cockiness" trait to such a degree that even a larger male was conspicuously subordinate.

When food was held out to the animals the aforementioned dominant male came forward rapidly to seize it, the other cats standing by. He did not growl as a cat so often does when it has a piece of food, *since no other cat approached him. After he was* eating, the others would come to take a piece of food quite indis*criminately, there apparently being no continuation of the social* hierarchy below the leader. If this dominant male was already satiated from previous eating, he allowed the others to come forward for a piece of food without protest. Dominance, therefore, is not specific to the feeding situation, but must have been determined beforehand. Later discussion will reveal how it takes place.

The question arose as to whether dominance-subordination behavior was related in any way to the usual masculine-feminine positions in sexual behavior, as Zuckerman and Maslow showed in the case of primates. Several times in the course of observations new cats were brought to the colony in groups of four or five. The *behavior of the dominant male was then striking. He strutted* and raised his back and tail in a more defiant manner than at any other time, and approached the new animals. He would then seize each in turn by the loose flesh at the nape of the neck and push its back down with his own hindquarters and mount as in copulation Rapid pelvic movements followed with erection. The dominant male would stand on the back of the prone subordinated animal to push him flat to the floor. The subordinate animal yowled and struggled, showing no sign of being a participant in copulation. It would dash away as soon as it could free itself from the dominating animal; jumping to a window sill or shelf where it cowered and remained *even hours at a time, fearing to come to get food at all. This reduc*tion of newcomers to subordination affected males and females alike. During several days the attacks were very frequent, and the yowling of the subjected animals continual. Great commotion prevailed in the colony until with amazing suddenness the attacks by the dominant male decreased in number, and appeared only sporadically thereafter. In this way each animal became subjected. The complete lack of the sexual component in the behavior of the subordinated animal was striking. This obviously precludes an explanation of the behavior in terms of sadism and masochism. The type of subjection just described occurred every time a new group was brought to live in the room, which was four times during the course of these observations.

Once the new animals showed complete submissiveness to the dominant male, quiet ruled thereafter in the colony. One could say that the dominance-submission relationship was now tacitly recognized by all. As it happened, no new animal brought to the room ever tried to dominate the acknowledged leader.

The dominant male became ill one day and was separated from the colony. It was a question then as to what would happen. For the first few days apparent communism in regard to eating and attitudes toward each other governed the colony. Apparently subordination had become habitual for all, including even the next most aggressive male. After the first few days, however, this next most aggressive male began to strut and show dominance in his general bearing as he moved around the room. Very seldom was he observed reducing the others to subordination by attacks. It would probably be correct to say that the others, already accustomed to subordination, acknowledged by their bearing their recognition of this male's superiority.

From our observations we would agree with Zuckerman and Maslow that the assumption of a masculine position for copulation by one cat and the feminine position by another, has no sexual significance, but that it fulfills a dominance-subordination purpose. Thus, homosexuality in the cat would be precluded. Cats, however, differ from both infra-human primates and roosters, it seems, in having no complete hierarchy of dominance and submission Passing down of punishment to a next lower animal in the subordination scale did not appear, as Maslow describes it in monkeys and apes and Murchison in roosters. Undoubtedly dominance and subordination are based more upon a physical basis than in monkeys and apes, where it seems peculiarly psychological, but not size alone, as we have shown, determines which one will be the dominant animal. Possibly such physiological tests as Murchison (7) used on the fowl would reveal a basis for it Dominance and subordination also appear in cats as well as in infra-human primates independent of the feeding situation. It is exceedingly doubtful that a female would ever play the dominating rôle

REFERENCES

1. MASLOW, A. H. The rôle of dominance in the social and sexual behavior of infra-human primates: I. Observations in the Vilas Park Zoo. *J. Genet. Psychol.*, 1936, **48**, 261-277.
2. ———. The rôle of dominance in the social and sexual behavior of infra-human primates: III. A theory of sexual behavior. *J. Genet. Psychol.*, 1936, **48**, 310-338.
3. ———. Individual psychology and the social behavior of animals. *Inter. J. Indiv. Psychol.*, 1936, 47-59.
4. MURCHISON, C. The experimental measurement of a social hierarchy in *Gallus domesticus*: I. The direct identification and direct measurement of Social Reflex No. 1 and Social Reflex No. 2. *J. Gen. Psychol.*, 1935, **12**, 3-39.
5. ———. The experimental measurement of a social hierarchy in *Gallus domesticus*: II. The identification and inferential measurement of Social Reflex No. 1 and Social Reflex No. 2 by means of social discrimination. *J. Soc. Psychol.*, 1935, **6**, 3-30.
6. ———. The experimental measurement of a social hierarchy in *Gallus domesticus*: III. The direct and inferential measurement of Social Reflex No. 3. *J. Genet. Psychol.*, 1935, **46**, 76-102.
7. MURCHISON, C., POMERAT, C. M., & ZARROW, M. X. The experimental measurement of a social hierarchy in *Gallus domesticus*: V. The post-mortem measurement of anatomical features. *J. Soc. Psychol.*, 1935, **6**, 172-181.
8. ZUCKERMAN, S. The Social Life of Monkeys and Apes. New York: Harcourt, Brace, 1932. Pp. 356.

Department of Psychology
Brooklyn College
Brooklyn, New York

The Journal of Genetic Psychology, 1933, 52, 429-432.

THE EFFECT OF EARLY REACHES ON HANDEDNESS IN THE RAT: A PRELIMINARY STUDY*

Department of Psychology University of New Mexico

Kenneth L. Wentworth

Introduction

Many studies have been made to find the cause or causes of handedness. In an unpublished study by Peterson (9), evidence has been found which appears to eliminate the factors of heredity, as he has failed to breed out left handedness after eleven generations. The theory of richness of blood flow to the left hemisphere being a cause of handedness has been discredited by Cunningham (3) who measured the carotids on either side and found no difference in size; by Peterson (7) who ligated the right carotids without increasing the incidence of right-handedness; and by Beeley (1) who pointed out that the anterior communicating artery connects the two cerebral arteries of the brain and forms the Circle of Willys, with the result that the cerebral blood supply is equalized The eye dominance theory seems to fail after Peterson (8) removed or enucleated right eyes of rats without altering the proportions of right- and left-handedness. It had also been noticed by Beeley (1) that the congenitally blind are also preponderantly right handed in about the same proportions as among the sighted The theory that the position of the foetus in the uterus may in some way be a controlling factor is weakened by Roos (10) who correlated foetal positions of 486 children with their subsequent handedness and who concluded that there was no causal relationship; and by Peterson (8) who removed one horn of the uteri of rats without changing the proportions.

Why should humans be preponderantly right handed and rats approximately 45 per cent right handed, 45 per cent left handed, with 10 per cent ambidexters? This is a discrepancy which none of the above mentioned factors (except possibly the last) could explain, even had they not been discredited as influences in the matter.

---

*Accepted for publication by C. J. Warden of the Editorial Board, and received in the Editorial Office on July 22, 1937

In at least one respect babies differ radically from rats. Babies receive hand training in varying amounts. Rats receive none. This training theory was first advanced by Sir Thomas Browne (2) in his *Vulgar Errors;* later by Humphry (4) and Kellog (5), and within modern times by Watson (12). An experiment by Watson with babies, showing their power to grip being about equal with either hand until the tenth day of age, led him to believe that it was not inherited, but learned. Schiller (11) claimed that the existence of left handed children caused his theory to be untenable. Peterson (8) found that ambidexterous rats having a greater preference for one hand do not seem to be fixated to an exclusive preference for that hand by a greater amount of practice with that hand. Opinion among the child psychologists as to the cause of handedness seems to favor maturation, with the knowledge that left-handedness in many children can be corrected by enforced practice.

Early Training for the Rat

Our problem here is to ascertain whether the proportions of right- and left-handedness may be altered by giving early right handed training to the rats.

Apparatus was set up which allowed the rats to get food only with their right hands (Figure 1).

Nineteen rats were placed in this apparatus, nine when 23 days old and ten when 35 days old. The younger group was allowed

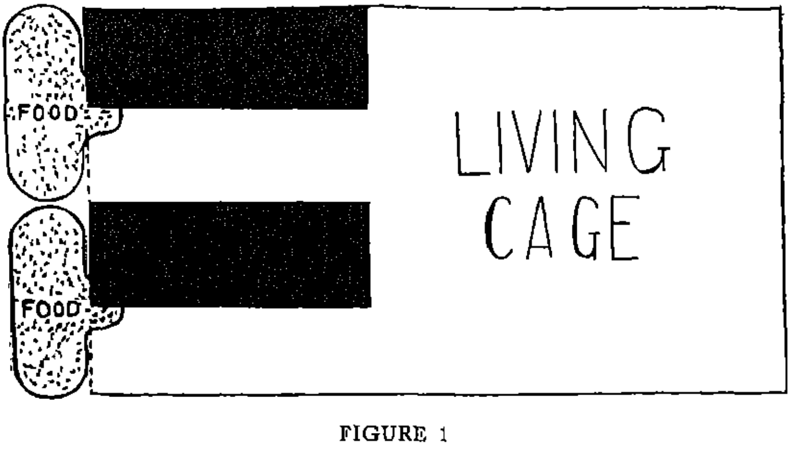


FIGURE 1

to remain 14 days, the older 21 days. During the training periods all their food was obtained by right handed reaches. Both groups were then returned to the cages for 90 days.

The younger group was allowed 14 days training against 21 for the older, so that the potency of fewer earlier reaches might be compared with more later reaches As one can see, the results of this experiment did not allow such a comparison to be made.

Final Tests and Results

After the 90-day rest period the rats were tested in the Peterson (6) reaching apparatus, in which food placed in a canary feeding dish could be reached equally well with either hand. Surprising results showed all 19 rats to be right handed over a period of seven days (Table 1). Peterson (8) considers this period adequate to judge consistency of reaches. Only two rats had any tendency toward ambidexterity, and this was slight (12 and 11 left handed reaches out of 175). To check the reliability of using a seven-day test, the rat which took 12 left handed reaches was tested 15 days. Its number of left handed reaches out of each 25 follows: 1, 0, 0,

TABLE 1

				Day				
Rat	1	2	3	4	5	6	7	
♂2	25	25	25	25	25	25	25	175
♂6	25	25	25	25	25	25	25	175
♂11	25	25	25	25	25	25	25	175
♂101	25	25	25	25	25	25	25	175
♂200	25	25	25	25	25	25	25	175
♂202	25	25	25	25	25	25	25	175
♀1	24	25	25	22	22	20	25	163
♀3	25	25	25	25	25	25	25	175
♀5	25	25	25	25	25	25	25	175
♀7	25	25	25	25	25	25	25	175
♀10	24	25	25	25	25	25	25	174
♀12	24	24	25	25	25	25	25	173
♀13	25	25	25	25	25	25	25	175
♀100	25	25	25	25	25	25	25	175
♀102	25	25	25	25	25	25	25	175
♀201	25	25	25	25	25	25	25	175
♀203	25	25	25	25	25	25	25	175
♀204	24	25	23	24	24	22	22	164
♀205	25	25	25	25	25	25	25	175

No right-handed reaches (out of 25) (out of 175)

3, 3, 5, 0, 2, 3, 1, 2, 2, 4, 5, 3. Fifteen of the 19 rats took no left handed reaches at all.

It would seem that forced early unilateral reaching in the rat has a real effect upon the hand used in later reaches.

Proposed Further Research

The writer is continuing the experiment, checking the effect of fewer days training, the effect of earlier and later training, the effect of left handed reaches directly at the end of the rest period and before the final testing period, the effect of left handed reaches after the final testing period, the effect of longer rest periods, the effect of varying both right and left handed reaches directly after the weaning period, the amount of transfer to other reaching situations (latch box).

References

1. BEELEY, A. L. The problem of handedness. *Trans. Utah Acad. Sci.* (1918-1921), **2**, 59-68
2. BROWNE, SIR THOMAS. Vulgar Errors. 17th century.
3. CUNNINGHAM, D. J. Left Handedness. Huxley Memorial Lecture, Anthropological Inst. of Gt. Britain, 1902.
4. HUMPHRY, G. M. The Human Foot and the Human Hand. 1861.
5. KELLOGG, G. M. The psychology of right- and left-handedness. *J. Amer. Med. Asso.*, 1898, **30**, 356.
6. PETERSON, G. M. A preliminary report on right- and left-handedness in the rat. *J. Comp. Psychol.*, 1931, **12**, 243-250.
7. ———. The influence of ligating one of the common carotid arteries upon handedness in the rat. *J. Comp. Psychol.*, 1932, **14**, 1-6
8. ——— Mechanisms of handedness in the rat. *Comp. Psychol. Monog.*, 1934, **9**, No. 6. Pp. 67.
9. ———. (Unpublished study.).
10. ROSE, M. M. A study of some factors entering into the determination of handedness. *Child Devel.*, 1935, **6**, 91-97.
11. SCHILLER, A. Theories of handedness. *J. Appl. Psychol.*, 1935, **19**, 694-703
12. WATSON, J. B., & WATSON, R. R. Studies in infant psychol. *Sci. Mo.*, 1921, **13**, 493-515.

181 Willard Street
Berlin, New Hampshire

The Journal of Genetic Psychology, 1938, 52, 433-437

ACHIEVEMENT AND INTELLIGENCE EXAMINATIONS CORRELATED WITH EACH OTHER AND WITH TEACHER'S RANKINGS*

Mooseheart Laboratory for Child Research

HAROLD A. KOHN[1]

As part of the regular routine work of this Laboratory, standard group intelligence and achievement examinations are given to the entire Mooseheart population annually. The general purpose of the achievement examination is for classification and guidance. The intelligence examinations have the general aim of selection and classification of children in terms of mental ability.

At the time of high school graduation the graduating class is very carefully ranked for academic standing; this ranking is inclusive of nothing but academic grades. Academic grades include grades in such portions of the curriculum as orchestra, physical and vocational work as well as the regular high school subjects. The intelligence examinations administered to the group comprise: Morgan's *Mental Test*, *Ohio State Psychological Examination*, and American Council on Education *Psychological Examination for High School Students*. The current achievement examination in use is the Sones-Harry *High School Achievement Test*.

The purpose of the study may be stated as (*a*) to determine the correlation existent between class standing and general intelligence, (*b*) class standing and the total achievement score, (*c*) to ascertain the relationship between achievement and intelligence, (*d*) to determine the intercorrelations of varied group measures of general intelligence.

Pintner (1) has compiled the correlations reported by 14 authors between general intelligence test scores and school marks made by high school pupils. These correlations vary from .28 to 60, the

---

*Recommended for publication by Martin L. Reymert, and received in the Editorial Office on August 11, 1937.

[1] Acknowledgment is made to Dr. M. L. Reymert, Director of the Mooseheart Laboratory, for having suggested this problem, and to Miss Ruth L. Thompson, Clinical Psychologist, for having very materially assisted in the preparation of the paper.

mean being .46. Mather (2), Hertzberg (3), Holzinger (4), and Hutcheon (5) have made studies dealing particularly with the reliability and validity of the Sones-Harry *Achievement Examination.* The results of these studies pertinent to the discussion here are: (*a*) correlation coefficients between percentile ranking on the test and teacher's grades range from .42 to .68; (*b*) correlations with Regents averages .57; (*c*) .75 with the Ohio State Psychological Examination, (*d*) the correlation between the American Council on Education Psychological Examination and the Sones-Harry is estimated by Holzinger at .80. Gates (6) worked with the problem of general intelligence in relation to school achievement, making use of elementary school pupils. The correlations of the group-test scores with a composite of educational achievement tests ranged from .47 to .65 with an average of .54.

The Sones-Harry *Achievement Test* scores (medians) are presented together with the norms in Table 1.

TABLE 1

	Grade 9	Norm	Grade 10	Norm	Grade 11	Norm	Grade 12	Norm
Lang Lit	35.4	36	37.2	44	48.5	53	40.4	60
Math	18.7	19	28.4	22	26.0	25	24.7	27
Nat Sci.	18.3	18	24.8	25	30.0	28	30.2	32
Soc St.	27.2	22	28.7	26	31.7	35	35.2	41
Total Score	99.0	92	121.50	117	136.0	139	141.7	163
N	70		53		78		55	

It is clear from Table 1 that while the ninth and tenth grades are, on the whole, doing slightly better work than may be expected of them, considering the Sones-Harry norms, the eleventh grade is working slightly below level, while the twelfth grade is definitely functioning below level[2]

When the results (raw scores) of a group test of general intelligence (Morgan's *Mental Test*) are correlated with the raw scores on the Sones-Harry *Achievement Test* we find that approximately the same correlations are found for the twelfth grade as for the other grades of the school system. The differences between grades

---

[2] A possible explanation of these results may be found from an inspection of the data below:

MEDIAN MENTAL AGE-MORGAN'S MENTAL TEST

9th Gr 15.1	10th Gr. 15.1	11th Gr 15.4	12th Gr. 16.1

TABLE 2

Morgan's Mental Test Scores Correlated with Sones-Harry Scores

	Grade	Sones-Harry	PE	N
	9	64	.05	68
Morgan's M	10	43	.08	46
	11	54	.05	76
	12	51	.07	52
			Total	242

in the zero order correlations reported in Table 2 are not statistically significant.

The intelligence examinations were all made within a few days of the Sones-Harry administration. The mean correlation between intelligence and achievement by other investigators has been about .45. The mean of the correlations above reported is 54, indicating when the results of Table 1 are taken into account that the school system is in general making the most of the student intelligence material it is working with, and that the correlation between intelligence and achievement appears to be higher at Mooseheart than over the country as a whole, as reported by previous investigations in this field.

The teacher's rankings twelfth grade were correlated with the Sones-Harry and with the average of the Ohio State and American Councils scores. The American Council on Education *Psychological Examination for High School* and *Ohio State Psychological Examination* were also correlated separately with the Sones-Harry scores These data are presented in Table 3.

TABLE 3

Correlations between Teacher's Rankings, Sones-Harry and Intelligence

	Sones-Harry	N	PE	Av. O S. U. and A C	N	PE
Teacher's Rankings	61	43	06	.59	43	07
American Council Psychological Exam	71	55	.04			
Ohio State Psych. Exam	82	42	03			

The mean correlation reported in the literature between intelligence and school grades is .46 as contrasted with the above correlation of .59 It is interesting to note that while the Morgan *Mental*

Test in the twelfth grade correlates only 51 with the Sones-Harry, much higher correlations are found between the American Council, and Sones-Harry, and the Ohio State and Sones-Harry. This is strongly suggestive of some common factors between the last two mentioned tests and the Sones-Harry Achievement Test, which are not present in the Morgan *Mental Test*

When the 43 seniors for whom intelligence, achievement and class standing scores are available, were ranked in order from 1 to 43 on each of the three measures, and these results graphically portrayed, a common phenomenon occurs Approximately 10 per cent of the distribution at each of the two extremes are found to closely parallel each other while the middle 80 per cent diverge to a much greater extent That is to say, in the cases of the very bright and the very dull students the achievement and class standing scores are in very close agreement. The dull make higher achievement scores in proportion to their intelligence than do the bright.

The zero order intercorrelations of all the group tests of intelligence reported in this paper, the Kuhlman-Anderson and the individual Stanford-Binet (old form) were computed (Table 4).

TABLE 4

Intercorrelations of Varied Measures of Intelligence

	K A	Morgan	Amer Coun.	Stanford-Binet	Ohio State	N
N Kuhlman-Anderson	(77) .71	(258) .78	—	(44) 69	(84) .72	463
Morgan Mental	78	(77) 76	(63) 92	(43) 71	(46) .67	229
American Council	—	.92	—	(64) .67	(47) .66	111
Stanford-Binet	69	71	67	(273) 81	67	273
Ohio State	.72	.67	.66	(68) 67	—	68
			Total number of cases			1,144

It is to be noted that in general the group tests of intelligence reported above correlate more highly with each other, than with the validating criterion, the individual Stanford-Binet examination.

SUMMARY

The mean correlation for all grades of the Mooseheart High School, between school achievement as measured by the Sones-Harry and general intelligence as measured by the Morgan *Mental Test* is .54. This is somewhat higher than previously reported correlations between these factors. The *Ohio State Psychological Examination* and the American Council on Education correlate .82 and 71 respectively with Sones-Harry, suggesting a possible common factor between these series of examinations, not present in Morgan's *Test*.

The teacher's ranking as expressed by the class standing, correlates .61 with Sones-Harry, and .59 with intelligence as expressed by the average of the Ohio State and American Council scores When class standing, intelligence, and achievement are compared, it is interesting to note that the dull student makes higher achievement scores in proportion to his intelligence, than does the bright. A table of correlations between various group tests of intelligence and the individual Stanford-Binet *Examination* (old form) is also presented.

REFERENCES

1 PINTNER, R. Intelligence Testing, Methods and Results New York Holt, 1921.

2. MATHER, I. A Validity of the Sones-Harry High School Achievement Test, Form A, for Use in Oregon. Thesis for M.S, Graduate School of the University of Oregon (August, 1930)

3. HERTZBERG, O. E An Evaluation of Some Factors Relating to the Selection of Freshmen (Pamphlet).

4 REEVES, R. Admission and Retention of Students University of Chicago Survey, Vol V (1933) Part VI, *The Value of Standardized Tests for Predicting Undergraduate Scholarship*, written by K. J. Holzinger.

5 HUTCHTON, ELSIE M A statistical analysis of the Sones-Harry High School Achievement Test. M A Thesis, Western Reserve University (1932)

6 GATES, A I The correlations of achievement in school subjects with intelligence tests and other variables *J Educ Psychol.*, 1922, **12**, 129-139, 223-235, 277-285.

Mooseheart Laboratory for Child Research
Mooseheart, Illinois

The Journal of Genetic Psychology, 1938, 52, 439-442.

A CASE OF APPARENT IMITATION IN A MONKEY*

Henry Phipps Psychiatric Clinic and Carnegie Laboratory of Embryology, Johns Hopkins Medical School

JOSEPHINE BALL

A small rhesus monkey was seen in a pet shop[1] lapping milk like a cat instead of sucking it up, monkey fashion. Its cage companions were a cat and another young rhesus who did not lap Upon inquiry the shopkeeper gave the following history.

Pepper, the lapping monkey, had lived with his father and mother until, at the age of six months, he was severely mangled by his father. He was removed from the parental cage and nursed back to health by the shopkeeper. After some three months he was as well as he ever would be but crippled and badly stunted.

When he was about eleven months old he was supplied with a kitten for company and a few weeks later a two months old rhesus baby, *Ginger*, was added to this family. It was during the three following months that the pictures, Figures 1 and 2, were taken At the end of this time the kitten was removed and two months later Pepper died, having lapped liquids like his feline friend till the end.

The shopkeeper said the new method of drinking had been adopted by Pepper a few days after the kitten had come to live with him. During these days, he said, Pepper would sit close to the kitten and watch intently while she lapped her milk Figure 1 shows Pepper lapping

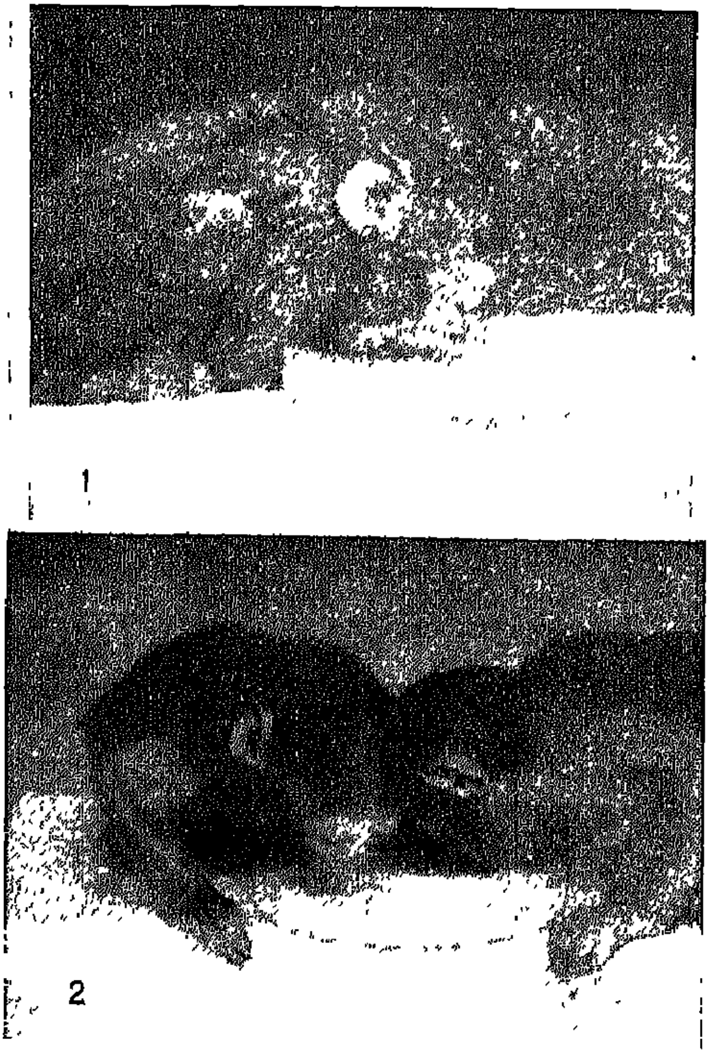
Ginger, much younger, never showed any tendency to lap Her method of drinking continued to be the monkey method of sucking as is shown in Figure 2 (animal on the right) where the cone raised on the surface of the milk by her sucking can be clearly seen. Ginger was never so friendly with the cat as was Pepper

Such a case of adoption of a new and apparently less efficient method of attaining a goal which had always been satisfactorily reached in another way is somewhat different from the cases reported

---

*Received in the Editorial Office on September 11, 1937.

[1] Baltimore Pet Shop, Mr. F. A. Snyder, proprietor.



FIGURES 1-2

(*Upper*) PEPPER LAPPING MILK

Compare with Figure 2, animal on the right.

(*Lower*) PEPPER (left) AND GINGER (right) DRINKING

Ginger's sucking method, commonly used by rhesus monkeys, is shown by the cone raised above the surface of the liquid.

of experimental studies of imitation. For a case of this kind the criteria demanded by Warden, Jenkins and Warner (2) as proof that learning had occurred through imitation and not by trial and error would seem unnecessary. The question here is not "Did he learn to reach the goal by observing or did he have to try out and discard movements that proved unsuccessful," for he already knew how to reach the goal. The question is rather "What induced him to discard a perfectly successful method?" Certainly it looks like what a child does when he cultivates a crooked smile because someone he admires has a crooked smile. Certainly emotional factors are implied in the history of this animal Can this change of method after opportunity for observation be interpreted in any other way than as imitation?

The decision here as to whether or not this is a case of imitation would seem to rest, for the skeptic, on the determination of the chances that Pepper might have hit upon the lapping method by accident in his original learning to drink.

Foley (1) describes the drinking of a yearling rhesus as a "combined process of sucking and lapping movements of the tongue" but these "lapping movements of the tongue" are the kind of tongue movements a human baby makes as part of the sucking pattern, the tongue moving in and out in cooperation with the sucking lips.[2] It is not the open mouthed, lipless lapping of a cat which is what Pepper did. His open mouth and retracted lips are well shown in Figure 1.

None of the workers in the Carnegie Laboratory of Embryology, where there have now been nearly 600 rhesus monkeys, has ever seen a monkey drink by lapping in this open mouthed fashion. The chance that Pepper had stumbled upon the method by accident before his association with the cat, unnoticed by Mr. Snyder, seems fairly small The case appears to add to the evidence recently presented by Warden and Jackson (3) that the rhesus monkey can imitate.

SUMMARY

A monkey, found living with a cat in a shop, drank by lapping like the cat instead of by sucking as monkeys normally drink The behavior is interpreted as due to imitation.

---

[2] Personal communication

REFERENCES

1. FOLEY, J P., JR Second year development of a rhesus monkey, (Macaca mulatta) reared in isolation during the first eighteen months. *J. Genet. Psychol*, 1935, 47, 73-97.
2. WARDEN, C. J., & JACKSON, T A. Imitative behavior in the rhesus monkey. *J. Genet. Psychol*, 1935, 46, 103-125.
3. WARDEN, C J, JENKINS, T N., & WARNER, L H An Introduction to Comparative Psychology New York Ronald, 1934 Pp. 581.

Carnegie Laboratory of Embryology
Johns Hopkins Medical School
Baltimore, Maryland

The Journal of Genetic Psychology, 1938, 52, 443-446.

THE STARTLE PATTERN IN INFANTS IN RESPONSE TO NON-AUDITORY STIMULI*

New York State Psychiatric Institute and Hospital

WILLIAM A. HUNT, FRANCES M. CLARKE, AND EDNA B. HUNT[1]

Previous work (2) has demonstrated the presence of the startle pattern in infants subjected to a sudden, intense auditory stimulus. This response is a generalized movement involving closing of the eyes, a characteristic facial expression, head movement, hunching of the shoulders, abduction of the upper arms, bending of the elbows, pronation of the lower arms, flexion of the fingers, forward movement of the trunk, contraction of the abdomen, and bending of the knees. In his original work in this field, Strauss (5) interpreted the response as an auditory reflex. The possibility immediately arises, however, that the pattern may be a response to any sudden, intense stimulus and not confined to auditory stimuli alone. This possibility is reinforced by the success of Hunt and Landis (3) in facilitating the response by a concomitant intense light stimulus. The general involvement of the entire body, plus the fact that the response is predominantly one of flexion, also suggests a protective function which might easily be of a generalized sort and not limited to auditory stimulation. This "protective" analogy seems particularly applicable in the large startle reactions shown in infrahuman primates (4). The present experiment was designed to bear upon the question of the generality of the startle pattern by subjecting a group of infants to sudden, intense stimulation of a non-auditory nature.

PROCEDURE

Nineteen infants, ranging in age from one to eight months, were used as subjects. Fifteen of them were male, and four female. The babies were photographed lying flat on their backs on a canvas cot.

---

*Received in the Editorial Office on October 1, 1937

[1] Thanks are due Miss Elizabeth Munger, Director, and the staff of the Connecticut State Farm at Niantic for their cooperation in furnishing facilities for this study.

The camera was placed above and was running at 64 frames per second (four times the normal speed). The stimuli were a sudden turning on of the ten photoflood photographic lights, a sudden puff of air in the face, a slap on the canvas cot on which the infants were lying, and a mild jab with a pin on the thigh. The films were then projected at slow motion and scored by a committee of three persons.

RESULTS

While startle patterns were infrequent following the non-auditory stimuli we used, there were cases in which they appeared At no time did the jab of the pin produce the startle pattern This may well have been due to the mildness of the stimulation, although several infants gave definite evidence through crying or random escape movements that they felt the stimulation Only one case gave a startle pattern following the puff of air, but this was definite and unmistakable. This infant was eight months old There were two cases, one at six weeks of age, and one at five months, who responded to the sudden turning up of the lights with the pattern. The slapping of the cot was the best stimulus we used. Startle patterns were elicited by this in five cases, one at six weeks of age, one at three months, two at four months, and one at eight months. In one more case at three months there was what appeared to be an incomplete response Another case at seven months showed a startle pattern complicated by movement which was taking place at the time of stimulation. Several times with the puff of air and the turning up of the lights isolated lid reflexes were observed, but in view of the specific nature of the two stimuli it seems best not to classify these as parts of an incomplete startle pattern.

DISCUSSION

While the appearance of startle in infants in response to our non-auditory stimuli is not frequent, it does occur. Our examples are not numerous, but are definite It seems necessary, therefore, to deny the claims of Strauss that this response is a purely auditory reflex, and to point out that it may be aroused by the sudden, intense stimulation of other modalities. The startle pattern is thus a more general response than Strauss had thought. This lends support to the view that it may be a primitive protective response of some sort. Further credibility is given this by the fact that the two best stimuli found to date, namely, a revolver shot (sudden loud noise) and slapping the canvas cot on which the baby is reclining (sudden loss of equilibrium) are well-known fear stimuli for infants. Any such attributing of a biological purpose to the response, however, involves no little inference and can at best be only a plausible hypothesis.

That the relative paucity of the startle patterns evoked by our stimuli was not due to the failure of the infant to feel the stimulation was shown by the appearance of random movement after the stimulation. The probability is that the present stimuli are not as intense as the revolver shot. Since intermodality intensity comparisons cannot be made except on a basis of amount of response, and since the amount of response is the question under consideration, we are caught in a hopeless circularity of reasoning. We are left with the fact that, whatever may be the reason, the stimuli used in this experiment are nowhere near as efficient as the revolver shot in calling forth the startle pattern.

It is of particular interest that the best of our stimuli should be the slap on the cot on which the infant is lying. This is a standard stimulus for the Moro reflex. That it should call forth both Moro and startle responses suggests some relation between the two. In his original study Strauss found the revolver shot producing both types of response. Since the Moro reflex has disappeared in the normal child by the age of three or four months, Strauss concluded that in very young infants the Moro reflex is the usual response to sudden intense auditory stimulation. As they develop, however, the Moro disappears and the startle response takes its place. Hunt, Clarke, and Hunt were chary of this assumption as they found two cases of overlap where the shot produced both startle and Moro response In such cases, the startle pattern comes in first with the Moro following it. Unless ultra-rapid photography is used it is difficult to see the fast startle response which often completes itself in less than half a second, and one tends to notice only the slower Moro reflex. In our present experiment the infant of six weeks who gave a startle response to the slap on the cot followed it immediately with a Moro reflex. This reinforces the finding of overlap in our previous work. Apparently both the Moro reflex and the startle pattern can be called forth by various sudden stimuli As the infant grows older the Moro reflex gradually disappears, meanwhile the startle pattern is developing and becoming more definite and stable, to persevere throughout life. There exists a period during which both reflexes are found in the infant and may be called forth by the same stimulus This has been discussed at greater length elsewhere (1).

SUMMARY

A group of 19 infants ranging in age from one to eight months were subjected to sudden non-auditory stimuli in an attempt to elicit the startle pattern A mild jab in the thigh with a pin did not produce any startle responses. A puff of air on the face produced a startle pattern in one infant. Two infants responded to the sudden turning on of the strong photographic lights. In seven of the cases where the cot on which the infant was lying was struck, startle patterns resulted. It is concluded that the startle pattern is not a mere auditory reflex but a general response to sudden, intense stimulation.

REFERENCES

1 GOLDSTEIN, K., LANDIS, C., HUNT, W. A., & CLARKE, F. M. The Moro reflex and the startle pattern. (In Preparation)

2 HUNT, W. A, CLARKE, F. M., & HUNT, E. B. Studies of the startle pattern. IV. Infants *J. of Psychol.*, 1936, 2, 339-352.

3 HUNT, W A., & LANDIS, C Studies of the startle pattern. II. Bodily pattern. *J. of Psychol*, 1936, 2, 207-213.

4 HUNT, W. A., LANDIS, C, & JACOBSEN, C. Studies of the startle pattern: V. Apes and monkeys *J. of Psychol.*, 1937, 3, 339-343.

5 STRAUSS, H. Das Zusammenschrecken. *J. f. Psychol u Neur.*, 1929, 39, 111-231.

New York State Psychiatric Institute and Hospital
722 West 168 Street
New York City

APPARATUS

The Journal of Genetic Psychology, 1938, 52, 447-448.

A COMBINED BLOCK PRINTING CELL AND TAPE HOLDER*

Cambridge, Massachusetts

Marie L. H Forbes

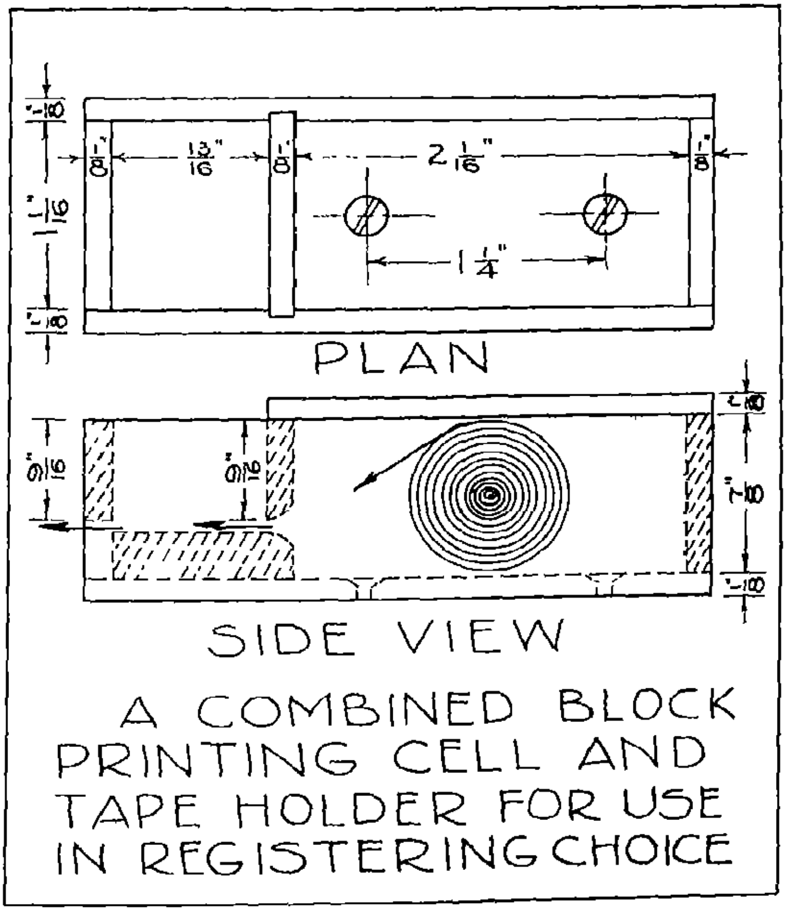


FIGURE 1

---

*Received in the Editorial Office on July 24, 1937

The drawing (Figure 1) shows a plan and side view of a box with two compartments in which tape unwinding from a roll in one compartment is pulled across the floor of the other compartment which serves as a printing cell.

Blocks to be used in registering choice are conveyed ready inked from a tray spread with inked material. The blocks, 1" x ¾" x ⅞", are separated by dowels ⅛" in diameter inserted in holes drilled in the sides of the tray at intervals of 15/16".

The writer gratefully acknowledges the examination of the device by her former instructor, Dr. Walter F. Dearborn, Director of the Psycho-Educational Clinic, Harvard University.

8½ Ash Street Place
Cambridge, Massachusetts

The Journal of Genetic Psychology, 1938, 52, 449-453

BOOKS

THESE JOURNALS WILL BUY BOOK REVIEWS

Our Old Policy. In the past we have published very few book reviews in these journals. That policy was based upon the conviction that the general run of scientific book review was trite trash that was not worth its space. That conviction has not been changed.

Reasons for a Shift in Policy. We have recently become interested in the problem of the fixation of professional opinion. Since 1920 we have witnessed the rise and decline of two great schools of psychology. Neither the rise nor the decline was based upon new discoveries in either case. The two series of phenomena were primarily lags in the fixation of professional opinion, and could have been avoided by the speeding up of free and open discussion. Wherever prestige lays its heavy hand, there is a lag in the fixation of professional opinion, and a new prestige must be built up to combat it. It would appear that a release of discussion would weaken the retarding power of prestige and hasten the fixation of professional opinion on particular issues.

Our New Policy. If a book is to be adequately discussed, every opinion concerning it must be given publicity, and recurring opinions must be registered as such. That is, a controversial book must be reviewed many times, and not just once or twice. We speak here of reviews at the professional level, not of childish blurbs or vindictive pecks.

Beginning immediately after the release of this notice in the *Journal of Genetic Psychology*, the *Journal of General Psychology*, and the *Journal of Social Psychology*, we will buy book reviews in the open market at not less than $1 per printed page and not more than $2 per printed page

Conditions. Only those books that are listed below in this section are eligible for such reviews All general elementary text-books are eliminated. Those books that are included, even though some of them might be poor, are probably the ones that make the most difference to psychology or deal with problems or techniques that are fundamental to psychology at the present time. The list begins with January, 1936, and the books are listed approximately in the order of receipt New 1938 books will be added as they come in. At the end of 1938, all of the 1936 books will be removed from the list.

A reviewer must possess the Ph.D. degree or its equal in training and experience. The Editor thinks of graduate students with affectionate regard, *but he cannot accept reviews of this type from* them.

A review must be written strictly at the professional level. It must not occupy less than two printed pages, nor more than twenty. It must avoid trivialities, such as chapter divisions, spelling of words, typographical errors, or any other matters that the reader is not looking for in that book A review must not be a soap-box or pulpit from which the reviewer exhibits his own intellectual hobbies or private feelings. It is the book that is being reviewed, and the book must occupy the stage. No competent review will use such make-believe weapons as "*but this is not psychology*," or "*but this is biology.*" The classifications of old-fashioned college catalogues are not of vital importance in these columns. A competent book deals with issues, or with techniques for the investigation of issues. A competent review identifies those issues, determines their importance, and evaluates the success or failure of the book in the accomplishment of its purpose. *For a reviewer to point out that the purpose of a book is not his* purpose, is in fact a statement by means of which the reviewer substitutes himself for the book. That type of vulgarity has no place on the stage of great books.

Procedure If among the books below there is one about which you have thought, and concerning which you have formulated some ideas, you are invited to write a review of that book. No matter if the book has been reviewed several times or a dozen times. It is important that your colleagues know your comprehension of the book and its significance In this way professional opinion will prevail quickly, and uncritical theories, unimportant or badly conceived investigations, misinterpretations, insufficient evidence, or uninspired work of any kind will stand revealed for what it is.

(Authors of reviews will receive a check immediately the size of which will be within the limits named above Nominations for inclusion in this list will be received gladly.)

(1936)

LEWIN, K. Principles of Topological Psychology New York: McGraw-Hill, 1936. Pp. 231.

Young, P. T. Motivation of Behavior. New York: Wiley, 1936. Pp. 562

Brown, J. F. Psychology and the Social Order New York McGraw-Hill, 1936. Pp. 529.

Terman, L M., & Miles, C. C Sex and Personality. New York: McGraw-Hill, 1936. Pp. 600

Guilford, J. P. Psychometric Methods New York McGraw-Hill, 1936 Pp. 566

Shaffer, L F. The Psychology of Adjustment. Boston Houghton Mifflin, 1936. Pp. 600

Warden, C. J., Jenkins, T. N., & Warner, L. H. Comparative Psychology: III. Vertebrates New York· Ronald Press, 1936. Pp 560.

Ruckmick, C. A. The Psychology of Feeling and Emotion New York: McGraw-Hill, 1936. Pp. 529

Washburn, M F. The Animal Mind. (Fourth edition.) New York: Macmillan, 1936. Pp. 536.

(1937)

Allport, G. W. Personality· A Psychological Interpretation New York· Holt, 1937 Pp. 588.

Beck, S J Introduction to the Rorschach Method *Res Monog., Amer Orthopsychiat. Asso.*, 1937, No 1.

Newman, H H, Freeman, F. N., & Holzinger, K. J. Twins: A Study of Heredity and Environment Chicago· Univ. Chicago Press, 1937. Pp 369.

Murphy, G, Murphy, L B, & Newcomb, T. M. Experimental Social Psychology. New York: Harper, 1937. Pp. 1121

Southall, J. P. C Introduction to Physiological Optics. New York and London· Oxford Univ. Press, 1937 Pp. 426

Terman, L M., & Merrill, M. Measuring Intelligence Boston Houghton Mifflin, 1937. Pp. 461

Thomas, W. I. Primitive Behavior· An Introduction to the Social Sciences New York: McGraw-Hill, 1937. Pp. 847.

Stagner, R. Psychology of Personality. New York· McGraw-Hill, 1937 Pp. 465.

Murphy, L. B Social Behavior and Child Personality New York: Columbia Univ. Press, 1937. Pp. 344.

Anastasi, A. Differential Psychology New York· Macmillan, 1937 Pp. 615

Burrow, T The Biology of Human Conflict New York. Macmillan, 1937. Pp. 435

(1938)

Thurstone, L. L. Primary Mental Abilities *Psychomet Monog.*, 1938, No. 1.

Spearman, C Psychology Down the Ages (2 vols) New York Macmillan, 1938. Pp 454 and 355

Gesell, A., & Thompson, H The Psychology of Early Growth. New York. Macmillan, 1938. Pp. 290.

Strong, E. K. Psychological Aspects of Business New York McGraw-Hill, 1938. Pp. 629

Boas, F The Mind of Primitive Man. (Rev. ed) New York: Macmillan, 1938 Pp 285

Sherman, M Mental Conflicts and Personality New York Longmans, 1938 Pp. 319.

BOOKS RECENTLY RECEIVED

Barnes, E. B, & Becker, H. Social Thought from Lore to Science (2 vols.) Boston: Heath, 1938 Pp. 790 and 400.

Bryan, R. C Pupil Rating of Secondary School Teachers New York: Teacher's Coll., Columbia Univ, 1938. Pp. 96.

Dexter, Emily S., & Omwake, Katherine T. An Introduction to the Fields of Psychology New York Prentice-Hall, 1938. Pp 236

Di Napoli, P. J. Homework in the New York City Elementary Schools New York. Teacher's Coll, Columbia Univ. 1938 Pp. 60

Gesell, A, & Thompson, Helen The Psychology of Early Growth. New York. Macmillan, 1938. Pp. 290

Grant, V. W. Psychological Optics. Chicago: Professional Press, 1938 Pp 240

Gray, W. H. Psychology of Elementary School Subjects. New York. Prentice-Hall, 1938 Pp 459.

Goldschmidt, R. Physiological Genetics New York. McGraw-Hill, 1938 Pp 375

Harrington, M. A Biological Approach to the Problem of Abnormal Behavior. Lancaster, Pa. Science Press, 1938 Pp 459

Hart, J K Mind in Transition. New York: Covici-Friede, 1938 Pp. 413

Hjort, J. The Human Value of Biology Cambridge, Mass. Harvard Univ Press, 1938 Pp 241.

Landis, P. H Three Iron Mining Towns. Ann Arbor, Mich : Edwards, 1938. Pp 148.

Margineanu, N. Analiza Factorilor Psihici. Cluj Univ Cluj, 1938. Pp. 199.

Margineanu, N. Elemente de Psihometrie. Cluj Univ. Cluj, 1938 Pp. 359.

Matisse, G. Le Primat du Phenomene Paris Alcan, 1938 Pp 280.

Mauge, F. L'Esprit et le Reel Percu. Paris: Alcan, 1938 Pp. 311.

Murphy, Lois B. Social Behavior and Child Personality New York Columbia Univ. Press, 1937. Pp. 344

Reinhardt, M R Social Psychology. Philadelphia Lippincott, 1938 Pp. 467.

Ruch, F. L Psychology and Life. Chicago. Scott, Foresman, 1938. Pp 679.

Spearman, C. Psychology Down the Ages. (2 vols) New York Macmillan, 1938. Pp 454 and 355

Sprott, W. J H. General Psychology New York· Longmans-Green, 1938 Pp. 446

Stern, W. General Psychology New York Macmillan, 1938. Pp 589

Strecker, A. S, & Chambers, F. T Alcohol. One Man's Meat. New York. Macmillan, 1938 Pp 230.

Tannenbaum, F. Crime and the Community. Boston Ginn, 1938. Pp 487.

Uhl, W L, & Powers, F. F. Personal and Social Adjustment. New York· Macmillan, 1938 Pp 475.

Vincent, Elizabeth L Mental Hygiene for Nurses. Philadelphia: Saunders, 1938 Pp. 263

Wallenstein, N Character and Personality of Children from Broken Homes. New York Teacher's Coll., Columbia Univ, 1938. Pp 86.